Artificial intelligence-2

A toy expert system

In *Computers & Society 5* we looked at some of the basic techniques and methods which are employed in the field of artificial intelligence (AI). In this article, we will look in more depth at **expert systems** and then go on to briefly examine **logic** as a problem solving paradigm, before surveying the current state of AI. We will then go on to look at the developments in computer hardware which will support large scale AI implementations.

In order to demonstrate the use of rule-based systems, let us look at a small 'toy' system which is designed to enable Robbie (the robot from the previous article) to spend a day at the zoo identifying the animals which he sees. To make the example a managable size, Robbie visits a very small zoo containing only seven animals.

The concept of **antecedent-consequent** rules was established in *Computers & Society 5*. This allows the truth of consequences to be established, or particular actions to be requested, under specified circumstances. Robbie needs a set of rules which will enable him, from what he can see of the animal, to deduce its name. First, we identify which biological class the animal occupies:

Rule 1	if	animal has hair
	then	it is a mammal
Rule 2	if	animal gives milk
	then	it is a mammal
Rule 3	if	animal has feathers
	then	it is a bird
Rule 4	if	animal flies
		it lays eggs
	then	it is a bird

Note that the last rule has two antecedent rules: this is because some mammals fly and some reptiles lay eggs, but only birds do both.

Now, if we know that the animal is a mammal, then we ought to find out whether or not it is carnivorous:

Rule 5	if	animal is a mammal
		it eats meat
	then	it is a carnivore
Rule 6	if	animal is a mammal
		it has pointed teeth
		it has claws
		its eyes point forward
	then	it is a carnivore

All the other mammals in this small zoo are ungulates:

Rule 7	if	animal is a mammal
		it has hoofs
	then	it is an ungulate
Rule 8	if	animal is a mammal
		it chews cud
	then	it is an ungulate
		it is even toed

Rule 8 demonstrates that we can have multiple consequent rules as well as antecedents.

We can now identify the particular mammals:

Rule 9	if	animal is a carnivore
		it has a tawny colour
		it has dark spots
	then	it is a cheetah
Rule 10	if	animal is a carnivore
		it has a tawny colour
		it has black stripes
	then	it is a tiger

Note here that the basic colour is redundant, since both are tawny, but there is no bar against superfluous information in the rule-base.

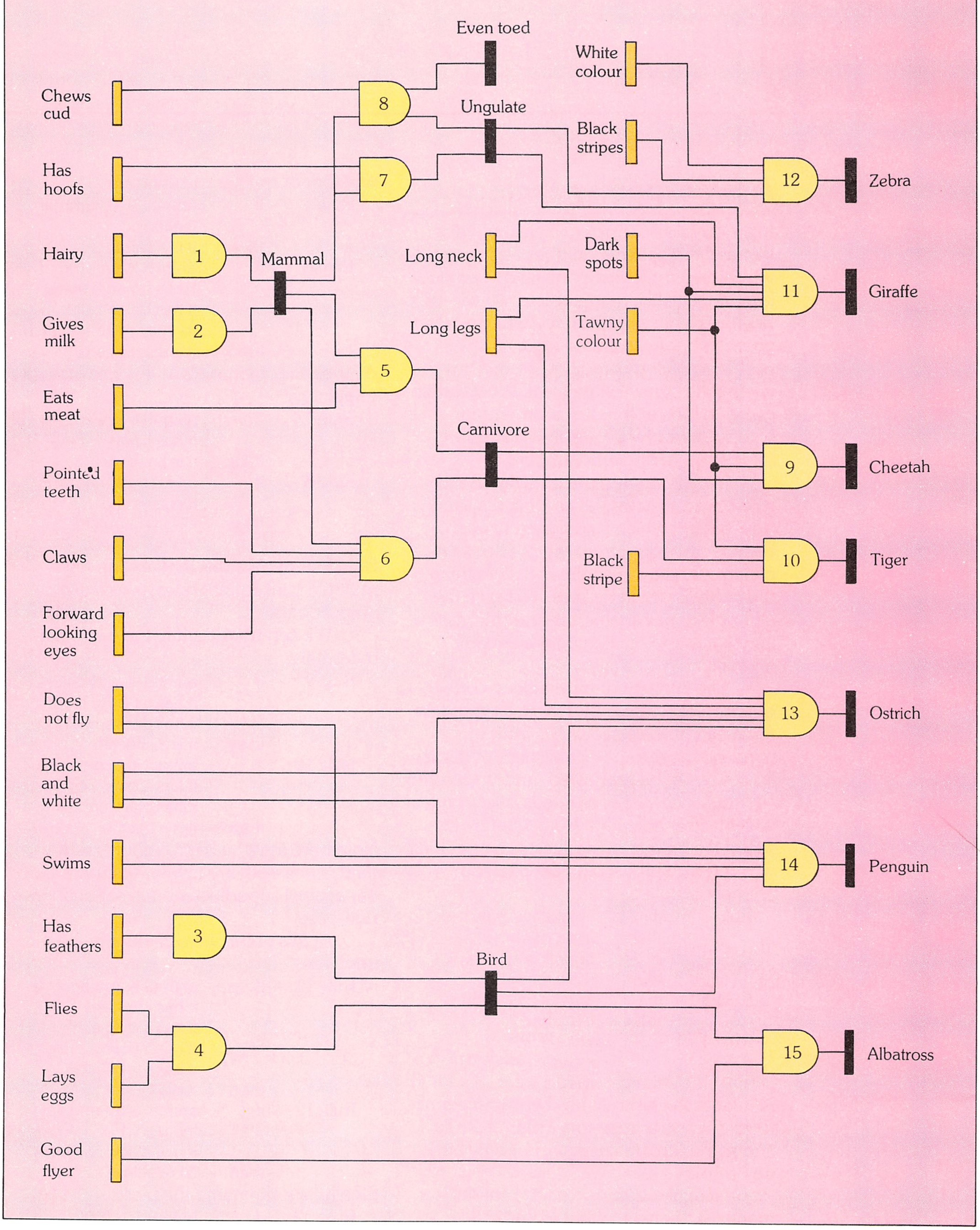
At this point, we can identify the ungulates:

Rule 11	if	animal is an ungulate
		it has long legs
		it has a long neck
		it has a tawny colour
		it has dark spots
	then	it is a giraffe
Rule 12	if	animal is an ungulate
		it has a white colour
		it has black stripes
	then	it is a zebra

Finally, we can name the birds:

Rule 13	if	animal is a bird

1


Even toed
White colour
Chews cud
8
Ungulate
Black stripes
Has hoofs
7
12
Zebra
Hairy
1
Mammal
Long neck
Dark spots
11
Giraffe
Gives milk
2
Long legs
Tawny colour
5
Eats meat
Carnivore
Pointed teeth
9
Cheetah
Claws
6
Black stripe
10
Tiger
Forward looking eyes
Does not fly
13
Ostrich
Black and white
Swims
14
Penguin
Has feathers
3
Bird
Flies
4
15
Albatross
Lays eggs
Good flyer

1. Inference network for the complete set of rules which will enable Robbie to deduce the names of the 7 animals in the zoo.

		it does not fly
		it has long legs
		it has a long neck
		it is black and white
	then	it is an ostrich
Rule 14	if	animal is a bird
		it does not fly
		it swims
		it is black and white
	then	it is a penguin
Rule 15	if	animal is a bird
		it is a good flyer
	then	it is an albatross

We can operate systems in both forward and backward-chaining modes. Remember that forward-chaining works from a series of facts towards a conclusion, and backward-chaining begins with a hypothesis and works backwards to find supporting evidence.

Forward-chaining system

Robbie sees a tawny animal with dark spots. After a while, the animal starts to feed a baby, and then settles down to chew cud. Therefore, the animal must give milk, and rule 2 triggers, and the deduced fact that the animal is a mammal is recorded. This deduced fact (mammal) and the observed fact (chews cud) come together and trigger rule 8, which reveals that the animal must be an ungulate and that it has two or four toes per foot. Robbie can see that the animal has a long neck and long legs, and this means that the antecedent conditions for rule 11 are all satisfied, thus the animal is a giraffe.

We saw the **inference network** for this part of the network of rules in the previous article. The inference network for the whole set of rules given above is shown in *figure 1*. As before, the orange rectangles represent raw or observed facts, and the black rectangles are the facts which are deduced by means of the rules, which are the 'AND gate' type structures.

Backward-chaining systems

Inference networks can be transformed into AND/OR trees, which were discussed in *Computers & Society 5*. The network in *figure 1* can be represented as shown in *figure 2*. Here, the top nodes of the tree are the resultant deduced facts shown at the right of *figure 1*. The filled circles are deducible facts, and the open circles represent the raw information which is needed to confirm a deduction. The AND nodes are formed by the rules, since all the antecedents must be satisfied before the consequence is true, and the OR nodes are formed by deduced facts which can be arrived at by more than one path.

For example, beginning with the hypothesis that the giraffe is, in fact, a cheetah, we have to satisfy the AND node 9. We first examine the animal and find that it has both a general tawny colour and dark spots, so two out of three of the inputs are true. Moving down a level, we have to see whether the deduction 'carnivore' is valid. This must be done by satisfying the OR node at this level. First we look to see whether it eats meat, and we find this invalid, so the AND node 5 is false. Turning to 6, we find neither pointed teeth, claws, nor forward-looking eyes, so this node is false. There are no other alternatives in the tree we have searched so far, so the hypothesis that the animal is a cheetah has been proven invalid. We could use the same procedure to prove that the animal is a giraffe.

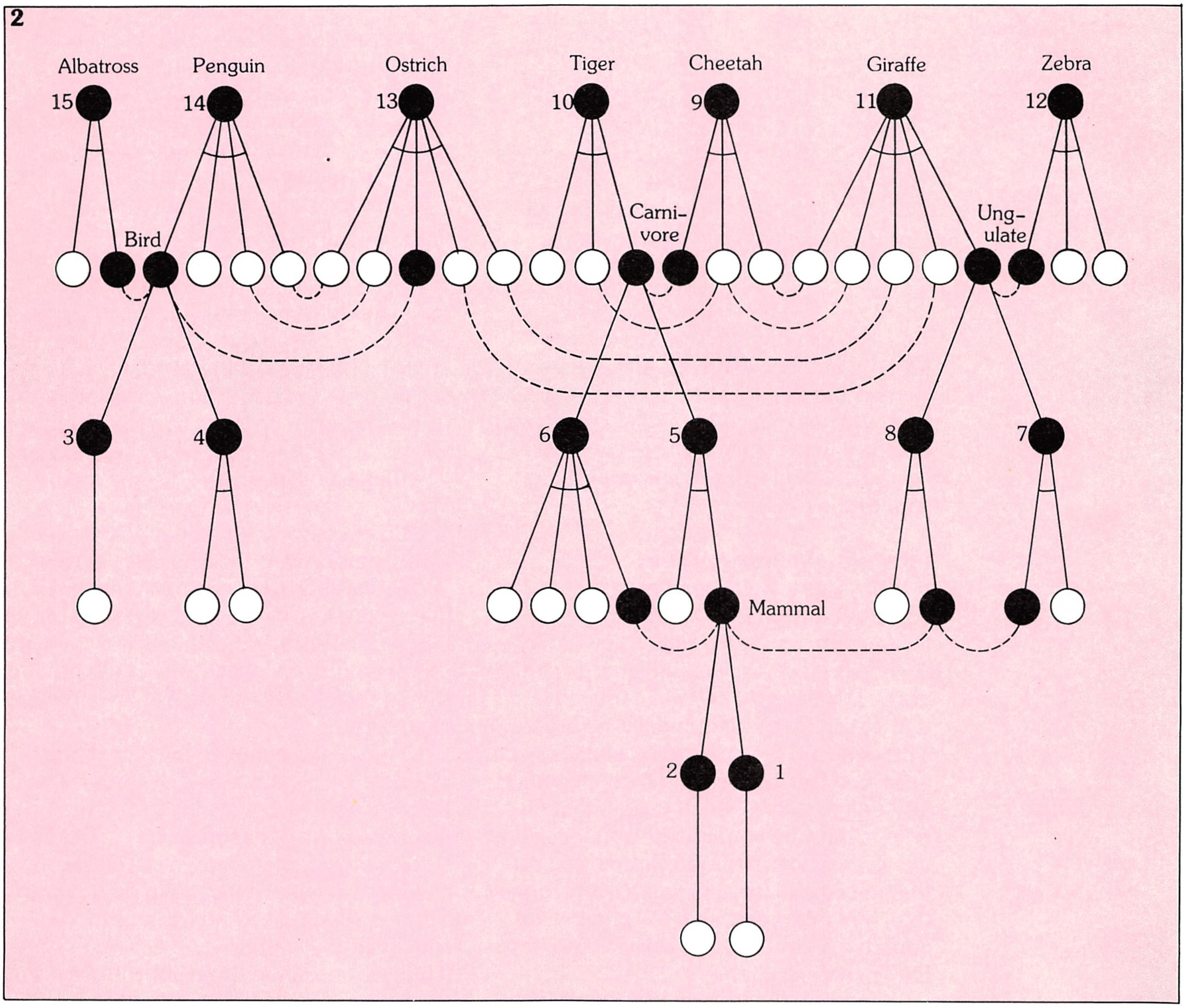
We have seen previously that one characteristic feature of AI systems is that they are able to *explain* their reasoning – this is also true of expert systems. The system must remember which rules it has used in arriving at a particular deduction. Suppose that rule 5 had been used to show that the animal was a cheetah. Then, if questions such as 'how did you show that the animal is a carnivore?' are asked, the answer can be determined by moving to the left and reporting that the animal is a mammal and that it eats meat. Ask 'why did you show it to be a mammal?' and the answer is given by moving to the right and reporting that it was necessary to show that the animal was a mammal so that rule 5 would confirm the animal as a carnivore.

Implementing an expert system

How is the **knowledge-base** generated? This is the role of the **knowledge engineer** and the **domain expert**.

The knowledge engineer is capable of understanding the software which is to be used to implement the system. This may either be written from scratch (1st generation systems); from a skeleton (adapted

2



2. The inference network of *figure 1* transformed into an AND/OR tree.

from previous work); or it might use a shell. Several shells exist, for example, Micro EXPERT, EXPERT, SAGE and AL/X, which are specially written systems consisting of knowledge-base construction aids, the conversational interface, and an inference engine (*figure 3*). An inference engine is the mechanism which performs the reasoning outlined above. The basic system contains an empty knowledge-base which must be built up for a particular application. Forward and backward-chaining systems are available.

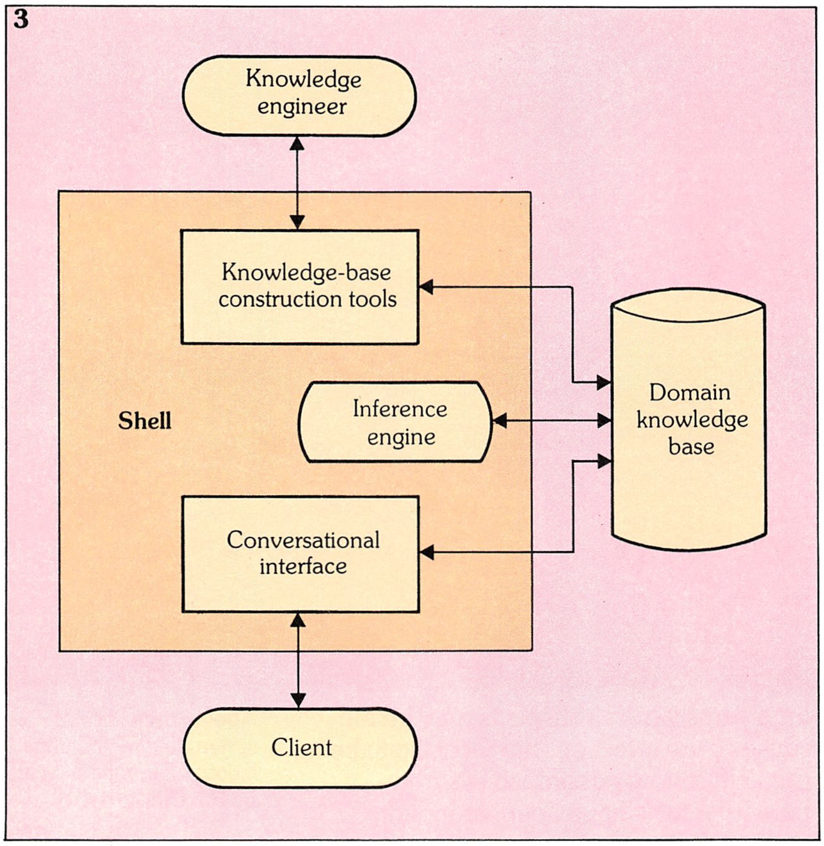
As you must realise, the expert has to know the **domain** of interest in great detail. Expert systems have been based on the expertise of physicians, geologists, chemists and many others. The recognisable characteristic of both the expert and the domain of the application is that the expertise can be formulated into a **set of rules**.

The knowledge engineer and the domain expert work together to translate the expert's knowledge into a coherent set of rules, which can then be efficiently used for inference within the selected problem solving structure. It is neither possible nor desirable to 'freeze' the knowledge about an application domain at a particular instant in time, nor to ensure that the system is 'correct' in all of its rules and rule relationships. Useful systems must therefore be under constant development. Thus, there are two interfaces to such a system, the **knowledge engineer's inter-**

face, allowing editing of existing rules, the addition of new rules etc., and the **client interface** which enables interaction with the system.

A terminal session with MYCIN

To give some indication of the power of this approach to the representation of knowledge, we will now look at the use of a more sophisticated system, MYCIN, which was developed to assist in the diagnosis of infective disease.



3. General structure of an Expert System, showing the boundary of a system shell.

Rules in MYCIN take the following form:

if:
1. the infection is meningitis, and
2. the patient has evidence of serious skin or soft tissue infection, and
3. organisms were not seen on the stain of the culture, and
4. the type of the infection is bacterial

then: there is evidence that the organisms which might be causing the infection are
staphylococcus-coag-pos (.75)
streptococcus-group-a (.5)

author: V. L. Yu

Literature: Robinson "Pneumococcal meningitis in sickle-cell anaemia" NEJM 274

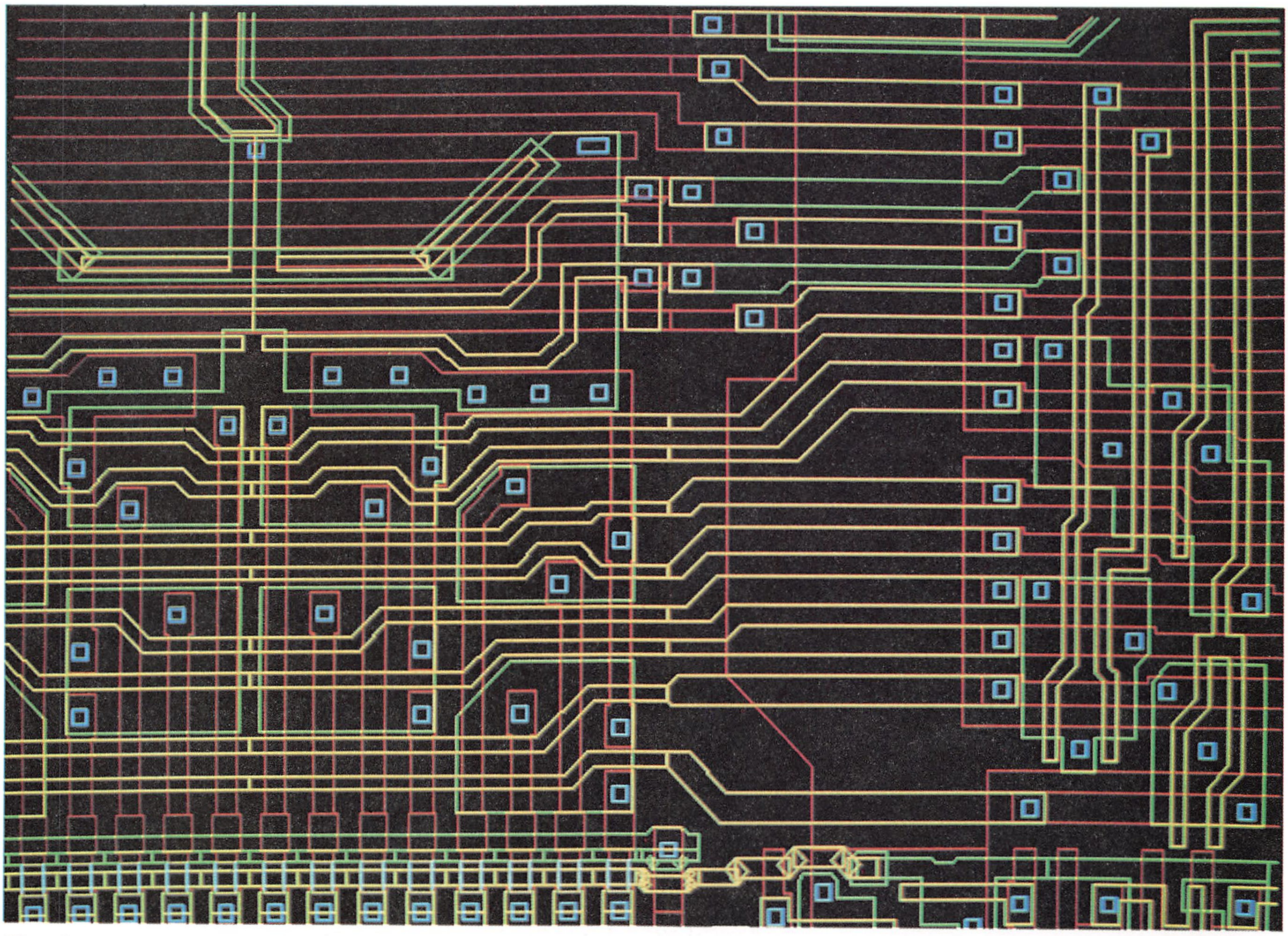
As you can see, the rules contain information about who entered them, and the basis upon which they were defined. Probabilities are assigned to the causal infections enabling the range of likely agents to be ordered, so that an appropriate treatment can be proposed.

How does a doctor question the system about a particular patient (say number 582)? Assuming that the initial information about age, sex, and some of the disease history has been completed, MYCIN asks question 17:

(17) Has patient 582 recently had symptoms of persistent headache or other abnormal neurological symptoms (dizziness, lethargy, etc.)?
** YES
(18) Has patient 582 recently had evidence of abnormal neurologic signs (nuchal rigidity, coma, seizures, etc.) documented by a doctor?
** *YES*

MYCIN can now narrow the field, and comments upon its current 'thinking':

The cultures will be considered to be associated with MENINGITIS.
(19) Please give the date of the first appearance of the symptoms.
** *24-Mar-84 21:30*
(20) Does patient 582 have any cutaneous lesions or rash on physical examination?
** *NO*
(21) Does patient 582 have evidence of ocular nerve dysfunction?
** *NO*
(22) Is patient 582 a compromised host (e.g. alcoholic, sickle-cell disease, immunosuppressed, etc.)?
** *YES*
(23) Is patient 582 immunosuppressed (e.g. corticosteroid therapy, cytotoxic drug therapy, radiation therapy, leukaemia, lymphoma, etc.)?
** *NO*
(24) For how many days had patient 582 had abnormal neurological signs?
** *6.3 HOURS*
•
•
•
(35) Is meningitis a hospital acquired infection?
** *WHY*



Above: using a CAD system to plan the track layout of a printed circuit board. (Photo: SGS).

The doctor wants to understand why it is important to know whether the infection was acquired while the patient was in hospital:

[2.0] This will aid in determining the organisms which might be causing the infection.

It has been established that:

[2.1] the infection is meningitis, and

[2.2] neurosurgery is not a relevant item in the history of the patient, and

•

•

•

[2.5] the age of the patient is greater than 10 days

Therefore, if

[2.6] the infection was acquired when the patient was hospitalised, then there is evidence that the organisms which might be causing the infection is *E.coli* (0.75), *staphylococcus-coag-pos* (0.3), *pseudomonas-aeruginosa* (0.3), *klebsiella-pneumoniae* (0.5)

MYCIN has given the probabilities of the organisms which it suspects may be the cause of the infection, the most likely being *E.coli*. It is now possible to ask how particular facts have been deduced:

** *HOW 2.5*

MYCIN then replies:

In answer to question 2 you said that the patient was 34 years old, therefore the patient is greater than 10 days old.

This question and answer dialogue can continue until the doctor is satisfied with the diagnosis that MYCIN has made, and then MYCIN can be asked to propose a treatment which will take into account the likelihood of particular organisms being responsible for the disease. For example, if the final diagnosis was that six different organisms were each equally likely, then a broad spectrum antibiotic might be suggested, whereas if MYCIN was certain that just one organism was responsible, then the antibiotic proposed would be specifically targeted at that organism.

State-of-art expert systems

Work on expert systems has been progressing for more than a decade now. Most of the research has been carried out by universities, but several commercial organisations are now using or selling expert systems. XCON for example, is used by a large manufacturer for computer systems design.

At this point we should look at some of the areas that are currently receiving attention in order to make greater use of expert systems. However, this summary excludes new application areas, as these are constantly expanding and changing.

One major problem is that expert systems can only cover a relatively small knowledge domain if they are to remain of manageable size. The reason for this is that the tools needed to maintain a large knowledge base are not yet available, so size tends to be self limiting.

Languages representing the knowledge of the expert also impose constraints upon what can be expressed. Once a framework has been fixed, people will always find knowledge to be represented which will not fit easily into that framework. This situation can be overcome by sophisticated programming, but this ad hoc approach is undesirable since it decreases standardisation.

At present, expertise has to be entered into the system through the medium of the knowledge engineer. If more sophisticated learning techniques were available, then this translation could be made more efficiently by the expert.

Both input/output languages and explanations of reasoning are too stylised; it is not yet possible to input free English text to the machine as it has only a limited vocabulary.

Another problem is that systems do not contain any information about their own assumptions or limitations. They can offer no guidance about whether the problem presented to them is truly outside or peripheral to their own expertise – they still try to apply the same rules. Socrates said (and we would be wise to remember it) that one of the marks of *wisdom* (as opposed to *knowledge*) was knowing when *not* to claim expertise.

For sophisticated systems, is is desirable that the knowledge base should contain the collected expertise of a group of experts. Unfortunately, the differences of interpretation between different experts raises a problem. Usually – unless they are blatantly contradictory – these differences are too subtle to be caught by the knowledge engineer or by a program. Therefore we must still rely on a single expert to arbitrate between conflicting views, and to dictate the final knowledge associations.

Logic and theorem proving

Formal logic is a relatively recent and very powerful addition to the available range of problem solving paradigms and, like other paradigms, it has both advantages and disadvantages. In its favour, the formal rules of logic, having matured over centuries, are clear, concise and well understood. However, it is easy to concentrate too heavily upon the the mathematics of logic, and divert attention from the powerful problem solving techniques available. The antecedent-consequent rules which we have already examined in some detail are really just a special case of logic. We will now look at some of the basic operations which are available:

1) **Predicates** are **functions** which map their arguments into TRUE or FALSE, e.g:

feathers(albatross)

is a TRUE expression. If:

feathers(x)

is a TRUE expression then the possible meanings of x are limited. The predicates known to be true are known as **axioms**.

2) We can combine predicates with the **logical** connectives AND, OR, NOT and IMPLIES, e.g:

bird(x) AND flies(x)

says that x is a bird which can fly:

NOT flies(x)

says that x cannot fly.

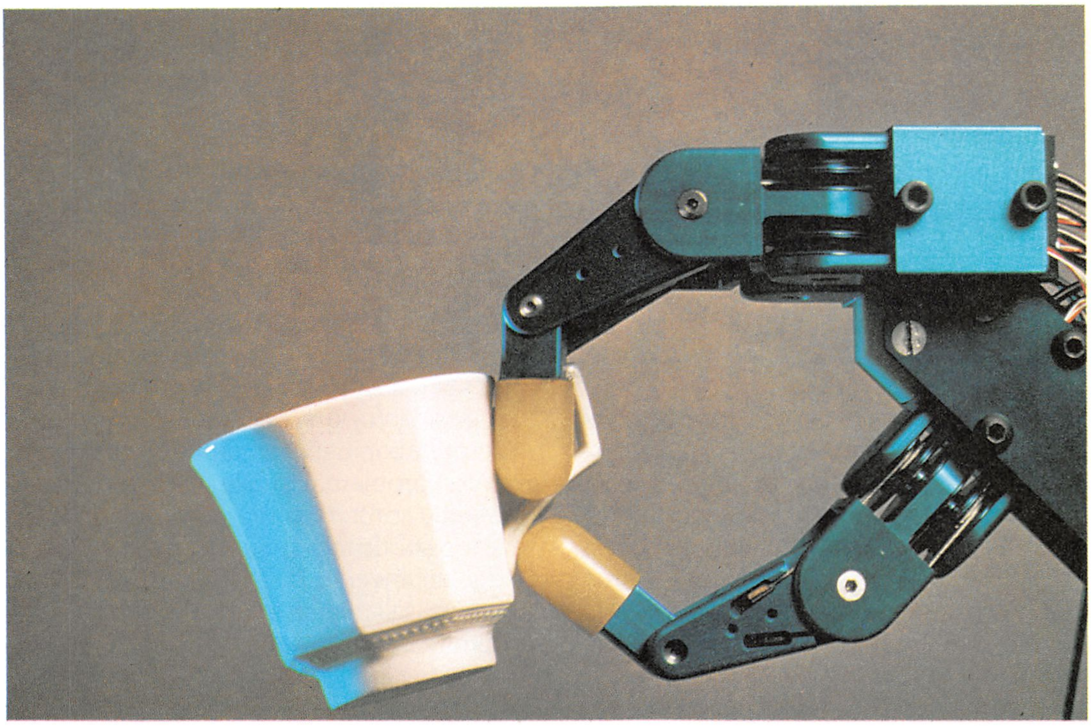
3) The AND operator forms **conjunctions**, the OR **disjoints** (and the parts are called **conjoints** and **disjuncts**).

4) We can also say that:

feathers(x) IMPLIES bird(x)

which resembles the antecedent-consequent rules we have already seen.

Consider an animal called 'squigs'. If:



Science Photo Library/Hank Morgan

Left: the 'Salisbury Hand' gripper robot under development at the University of Massachusetts.

feathers(squigs)
feathers(x) IMPLIES bird(x)

are axioms then we can begin to investigate **proof**. The rule of inference **modus ponens** says that if there are axioms of the form:

E1 IMPLIES E2
E1

then E2 logically follows. IF E2 is the objective of the proof (in this case 'squigs' is a bird) then we have finished. If not, then we could continue with this procedure, generating new axioms (for E2 is also an axiom after the above proof) until the correct deduction arises. Thus you can see that this gives us a means for deducting facts which were not anticipated by the knowledge engineer. In the rule-based approach, deducible facts are limited to those which form the 'consequence' part of the rules. Here, we can reason from raw facts and produce new facts which take into account all the axioms. The problem with modum ponens is that what we are attempting may not be TRUE, and we would then not know when to stop our search for the proof which does not exist.

Resolution is a better rule of inference. Resolution says that if there are axioms of the form:

E1 OR E2
(NOT E2) OR E3

then:

E1 OR E3

logically follows. To see this, consider E2 to be TRUE: then (NOT E2) is FALSE; if (NOT E2) is FALSE, then from the second expression, E3 must be TRUE; therefore E1 OR E3 must be TRUE. Alternatively, suppose E2 is FALSE; then from the first expression, E1 must be TRUE; then, E1 OR E3 must also be TRUE. So, providing both expressions are TRUE, the consequence follows.

The resolution proofs of theorems in real systems are based on the following rationale:

1) assume the negation of the theorem is TRUE;
2) show that the axioms and the assumed negation together determine something to be TRUE which is known *not* to be TRUE;
3) conclude that the negation cannot be TRUE since it leads to a contradiction;
4) conclude that the theorem must be true.

This is known as **proof by refutation**. For example, suppose that we have:

(NOT feathers(squigs)) OR bird (squigs)
feathers(squigs)

Satisfy yourself that this means that 'squigs' is a bird. Now add the negation of the theorem to be proved:

(NOT feathers(squigs)) OR bird(squigs)
feathers(squigs)
(NOT bird(squigs))

Resolve the first two expressions in the way given above to yield:

(NOT feathers(squigs), OR bird(squigs)
feathers(squigs)
(NOT bird(squigs))
bird(squigs)

Now we have a contradiction, 'squigs' cannot be both a bird and not a bird. Therefore the assumption which led to the contradiction was FALSE, and the theorem

bird(squigs)

must be TRUE.

A new generation of computers

We now turn from the mechanics of AI to a survey of what is happening in the AI world today. A great deal of interest and speculation has been aroused by the announcement by the Japanese that they are funding, to an enormous extent, the development of **fifth generation** computers. Why do we need a new generation of computers, and how are they different from the existing, fourth generation machines? *Table 1* summarises the characteristics of the five computer generations; performance is measured in 'ips' (instructions per second).

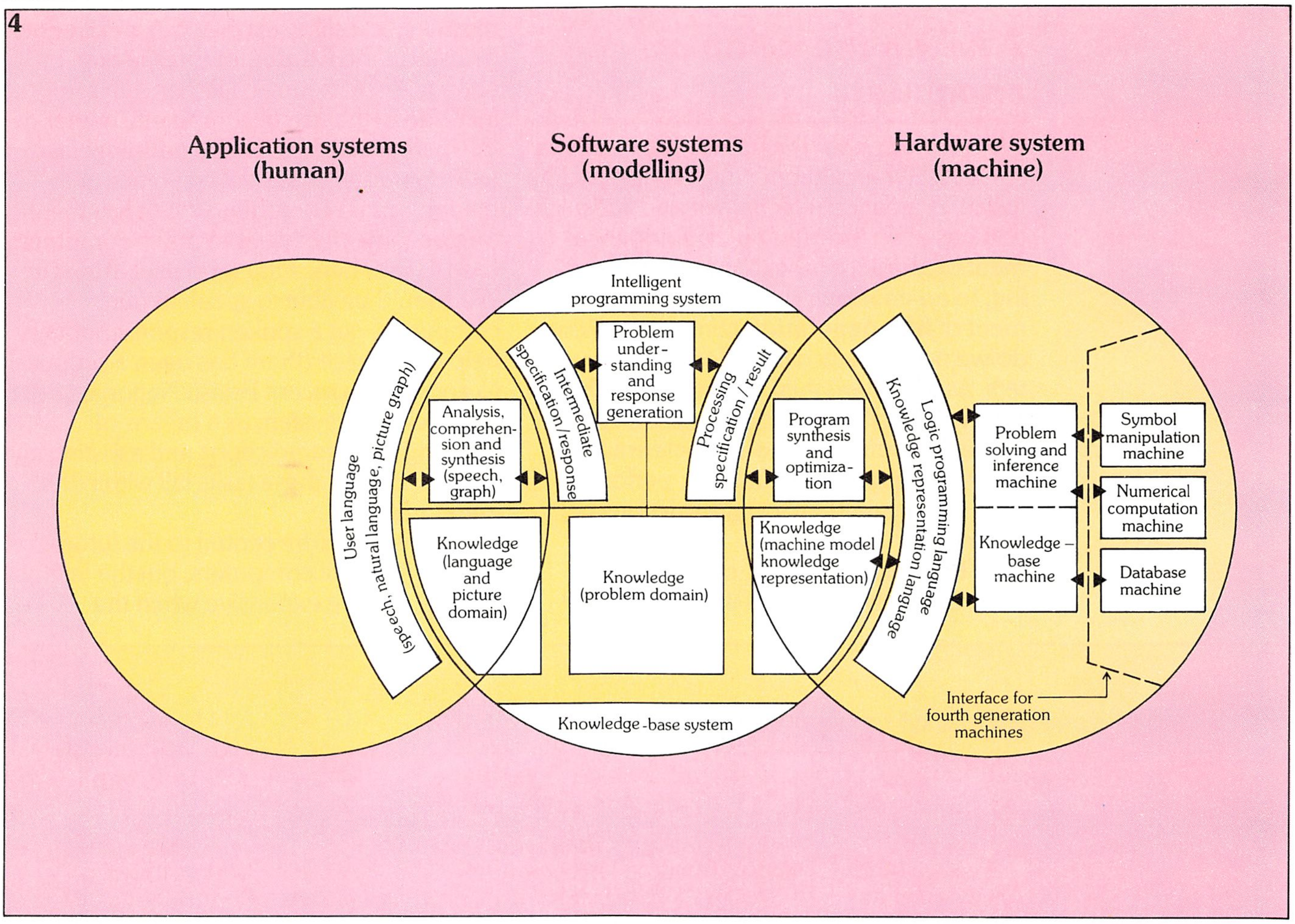
In today's societies, computers are essential to civilisation as we know it. We may like or dislike this fact, but we cannot gainsay it and we cannot foresee any change in the trend. Together with the increase in the use of computers, comes the increase in their use by **non-specialists**: computer aided design, office automation, home computing etc. It has therefore become necessary to make computers easy to use; hence the term **user-friendly**. We should be able to speak to our computers, get a spoken response, obtain help and advice (there has been a proposal to encode the recent British Nationality Act into an expert system as a means of showing up contradictions and assisting lay persons). All of these tasks fall within the ambit of AI.

AI is therefore central to the future development of computers, but the techniques it uses are expensive when the

Table 1
The five generations of computer technology

Criteria \ Period	First 1946-1956	Second 1957-1963	Third 1964-1981	Fourth 1982-1989	Fifth 1990 . . .
Hardware	Valves Magnetic Drum and Cathode Ray Tube Main Memory	Transistors Magnetic Core Memory	Integrated Circuits Semiconductor Memory Magnetic Disks Minicomputers Microprocessors	VLSI Bubble Memory Optical Disks Microcomputers Distributed Array Processors	Ultra Large Scale Integration 3D Circuit Design Parallel Architecture Gallium Arsenide Josephson Junctions Optical Components
Software	Machine Code Autocode Stored Programs	High-Level Languages: FORTRAN ALGOL COBOL	Advanced Languages Structured Programming: PASCAL; LISP Timesharing Computer Graphics	ADA Expert Systems Object-Oriented Languages PROLOG Packaged Systems	Concurrent Languages Functional Programming Symbolic Processing: Natural Language Vision Speech
Telecommunications	Telephone Teletype	Digital Transmission Pulse Code Modulation VDUs	Satellite Microwave Packet Switching Networks (LANs, WANs)	Integrated Systems Digital Network (ISDN) Optical	Modular Systems Cellular Radio Merged Communication and Computer Technology
Performance	10 kips 1 kbyte	200 kips 32 kbyte	5 Mips 2 Mbyte	50 Mips 8 Mbyte	1 Gips – 1 Tips ?

ips = instructions per second; lips = logical inferences per second; k = 10^3 (kilo); M = 10^6 (Mega); G = 10^9 (Giga); T = 10^{12} (Tera)



4. A conceptual diagram of a fifth generation computer system from the programmers' viewpoint.

knowledge base is of reasonable size. Even laboratory examples – small scale demonstrations – use enormous resources. Most present-day computers use the **von Neumann** architecture, i.e. one instruction is being executed at any instant. This serial architecture was developed when computers were expensive to manufacture. If several tasks could be performed simultaneously then, providing they were synchronised, the total task could be performed in much less time. The rapid developments in microelectronics have made the actual processor the cheapest part of many computers – it is more expensive to build, power, store, distribute and sell a personal computer than to buy the microprocessor at its heart. Therefore, the serial architecture can now be superseded by some form of parallel system, where processing elements co-operate in the solution of the total problem. The latest developments in this area, very large scale integration (VLSI), involve designing systems containing from one to five million transistors onto single chips of silicon six or seven millimetres square.

So the further development of computers can be seen to rest upon three main challenges:

1) **artificial intelligence** and the ability to make computers solve sophisticated intelligence problems;
2) **parallel computer architectures** and the ability to subdivide complex tasks into subgoals which smaller processors can solve independently: and
3) **microelectronics** which must continue to reduce the size and cost of parallel processors so that highly complex applications can be supported.

Figure 4 shows a schematic overview of the direction in which research is leading. The right-hand circle shows the hardware system which will implement the inference system directly. This will enable reasoning to take place at much higher speeds. Performance of the new machines

will be measured in **logical inferences per second** (lips). The target performance for the new machines is in the range 100 Mlips to 1 Glips. (1 lips is roughly equivalent to 100 ips to 1 kips, so an inference engine running at 1 Glips will have the power of a conventional machine running at 1 Tips.)

Currently, the largest fourth generation machines perform in the tens and hundreds of Mips range, so the performance increase over the next decade will be about 10,000 times. The central circle shows the software which will be necessary to develop and run the applications envisaged, and the left-hand circle designates the applications support area.

Competitive research in information technology

The Japanese New Generation computer project has stirred up a worldwide flurry of research activity. In 1979 the Japanese Ministry of International Trade and Industry (MITI) set out to envisage the information world of the 1990s. After consultations and working party reports, the Japanese announced their national project at the International Conference on Fifth Generation Computer Systems, held in Tokyo in 1981. The American and European scientists who were present, concluded that Japan was no longer willing to take western technology and improve upon it, but that they intended to grasp the initiative and develop unprecedented systems of their own. This single conference has resulted in dramatic responses in terms of the funding of this type of research.

The Japanese have established a central research centre, the Institute for New Generation Computer Technology (ICOT). This centre consists of 52 staff (who have proven research records) all under 35 years of age. The centre is equally funded by eight of the large Japanese corporations, who also share equally in the results – staff being seconded from these organisations. ICOT contracts with outside organisations (university laboratories, private companies, etc.) and engages in collaborative research, thus creating a large body of expertise. A large amount of government money is available to support the project, which is to be of ten years duration.

The Americans, unlike the Japanese, have funded research in AI for 20 years or more, and so possess a natural advantage. Nevertheless, there was no co-ordinated national programme other than that funded by the Defense Advanced Research Projects Agency (DARPA) until after the Japanese conference. DARPA is funding a major long-term research programme called Very High Speed Integrated Circuits (VHSIC) which is intended to unlock the physics and manufacture of very small, very fast devices.

Three new concurrent research programmes have now been founded: the Microelectronics and Computer Technology Corp. (MCC), the Microelectronics Center of North Carolina (MCNC), and the Semiconductor Research Co-operative (SRC). The first, located in Austin, Texas, is a direct counter to ICOT and its focus will be on packaging, CAD/CAM, software engineering and computer architecture. The other two organisations are really research brokers. MCNC distributes state grants, and SRC business cash, to particular research groups for specified work. The American objective is to retain their lead in information technology which is threatened by Japanese proposals.

The perception in Europe is of falling farther behind. With the objective of reversing the negative balance of trade and technology transfer, the European Economic Community has taken the initiative by funding the European Strategic Programme on Research in Information Technology (ESPRIT). This is similar to the Japanese and American programmes in that it involves precompetitive collaboration between different companies and research groups, but it has the added zest of being an international partnership. British companies are well represented in the ESPRIT programme, being involved in more than half the projects.

The only national programme in Europe has been established in Britain, though the French have been considering such a programme for several years. Known as the Alvey Programme, after the chairman of the committee whose report led to its foundation, this programme sets out to fund work in four major areas in information technology:

1) **Software engineering** is the development of new languages, the techniques for software development, and tools such as correctness verifiers, automatic programmers, etc.
2) **Intelligent knowledge-based systems** (IKBS) which is the AI part of the British programme.
3) **Man-machine interfaces** (MMI) is research into the ways that humans interact with computers: browsing through database structures; graphical and speech interfaces, etc.
4) **VLSI and CAD** focuses on research into the physics of small devices, how they can be manufactured, process technology, and computer tools for the management of designs which include millions of transistors.

The organisation of Alvey is consciously modelled on the Japanese MITI and the American DARPA, and includes the Ministry of Defence, the Department of Trade and Industry, and the Science and Engineering Research Council. Like ESPRIT, this is a precompetitive research programme, the four areas are identified as **enabling technologies**, and is intended to provide an environment in Britain which will result in commercial development. The programme runs for five years, and is funded 50% by government and 50% by business.

A day in the life of a fifth generation man

Before looking in any detail at the proposals for fifth generation computer hardware, let us speculate a little on one day in the life of a man living in the age of fifth generation machines, say around 1995. All of the techniques which will be mentioned have been demonstrated in research laboratories, only the specific machines and situations are fictional. Let us imagine a solicitor who lives in Reading and works in the City of London for a large multinational company. His name is Rob.

Rob is driving into work one morning along the motorway. Over the radio, he hears that there are carriageway repairs and a five mile tailback at Slough. He turns on his fifth generation terminal which, via **cellular radio**, connects itself over the voice-grade telephone system to his office workstation. With spoken commands (analysed by a system trained to respond to his voice) he requests information on alternate routes to his office together with their congestion. The office computer then connects itself over a **wide area network** to a computer centre which maintains optical disks on which road maps of the whole country are stored. The computer at this centre can retrieve maps of the appropriate areas and uses search strategies to find alternate routes, taking into account the traffic situations which are monitored via traffic flow sensors located along most major roads. The alternatives are passed back to the office computer, together with a bill for the service provided.

The alternatives are converted into speech by the office machine, and Rob listens to the summaries of the routes and the predicted traffic. He tells the terminal which he prefers, and is passed the next part of the route as he reports his current location.

Once the route is settled, he asks for his **electronic mail** to be read to him. One of the letters is urgent, so he immediately dictates a reply, which is compressed and has the grammar improved. This is sent off to the recipient with copies to the people he names. Another letter requests him to fly to New York for a conference with American colleagues. Although teleconferencing is used for most conferences, face-to-face interaction cannot be improved upon. He asks the computer to book his flights and a rented car for the specified dates, but decides to wait before booking the hotel so that he can look at a 'book' on New York which he asks to be obtained from the appropriate agency. Once his diary has been checked, the office computer invokes a trip-planning expert system to find the best routes and prices. After the personality profile for Rob has been consulted, the central airline reservations computer is informed, and the system asks for a window seat and for vegetarian meals. The computer of a car rental firm is told of his requirements, ordering a large car because of the entertaining that Rob will have to do. After completing this business, Rob asks the system to play him the Bach Double Violin Concerto, and the terminal finds the piece on its optical disk and reproduces it digital-

ly while he completes his drive to the office.

Once at the office, Rob wants to book his hotel. His workstation has high-resolution colour graphics and is equipped with a touch sensitive display. He uses this to browse through the book on New York which has been sent via satellite to his office computer in response to his earlier request. He asks for a sequence of images near the Rockerfeller Centre which is where his meeting is scheduled. The scene looks like a video tape of the city, but Rob can control what images he sees by touching left or right-turn controls on the screen which gives the sensation of actually walking in the city. He can enter a hotel, and examine its rooms, until he finds one that he likes, when he can instruct the trip-planner to make the reservations.

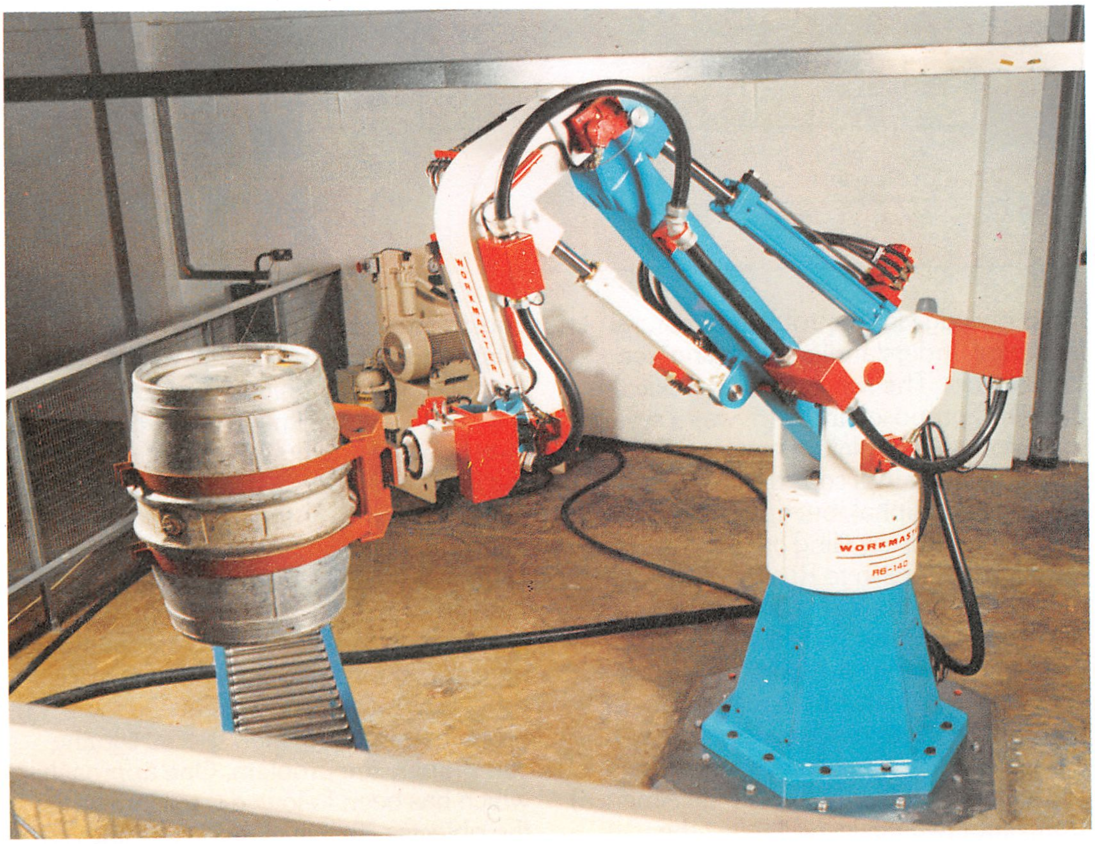
He can now request expert system help in preparing the proposal which he needs for the meeting. He can generate the report and file it, and then distribute it to other members of the group for comments. Finally, before going home, he can relax by wandering around the streets of Istanbul!

The future of AI

We now have a sketchy image of the sort of intelligence that future computer systems will have. What, then, are the major research topics currently going forward in AI?

Probably the biggest current bottleneck preventing AI progress is the problem of knowledge acquisition. It was noted earlier that present-day expert systems involve a knowledge engineer painstakingly translating the expertise of the domain expert into machine readable form. The largest knowledge bases which have so far been generated consist of between 500 and 1000 rules and have taken a decade or so to develop to their present state. To be ready for 1992 (the end of the Japanese programme) then, we should have started to enter the 10,000 or so rules which are estimated to be necessary for the projected applications in 1892! Clearly, some means of faster, automatic, acquisition of knowledge must be developed.

Several approaches are being followed. We touched on one of them in our earlier consideration of state of the art



Right: the movement of the Workmaster industrial robot is controlled by seven microprocessors. It is able to 'soft place' heavy weights and co-ordinate its movement with that of a conveyor belt. (Photo: Workmaster).

expert systems: better descriptive languages and knowledge representations so that the expert himself can transfer the expertise. Much work is also being directed at the problem of **learning**. Over centuries, civilisation has patiently documented its knowledge in the form of books. So how can machines be taught to *read* books, to convert text into rule structures with which they can reason, resolve contradictions between different texts/experts, and remove knowledge which is subsequently shown to be erroneous?

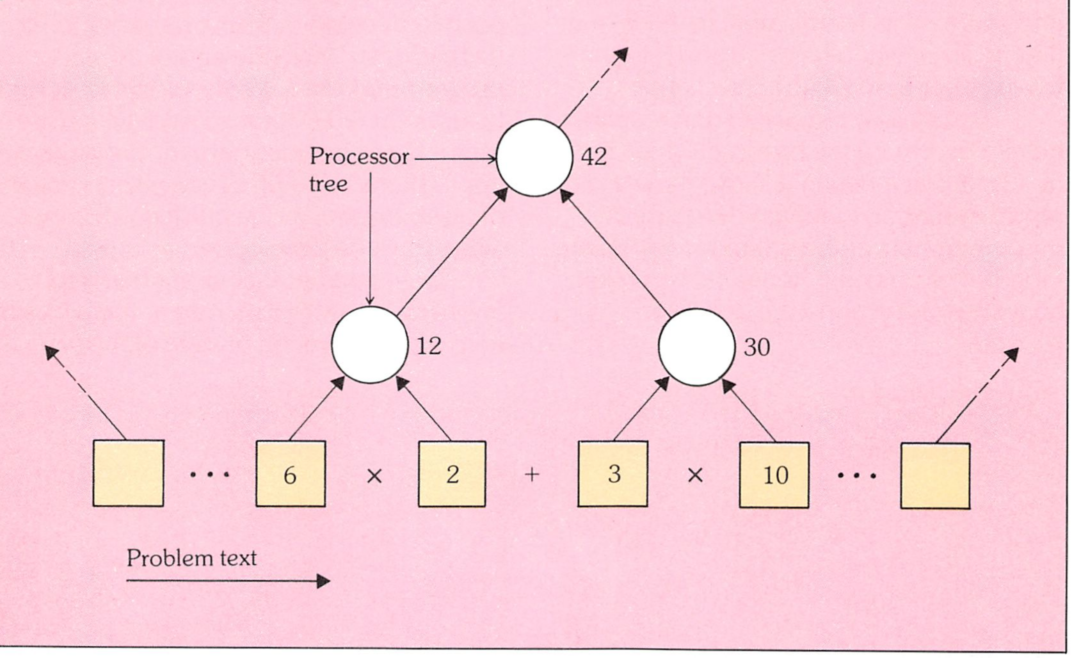
One problem which needs to be solved before it is possible to 'read' books for knowledge is the problem of natural language understanding – reading free English text. The first system to tackle this problem, proposed in 1946, was a system for automatic translation between English and Russian. It was considered a failure: it may be apocryphal, but apparently translating 'the spirit is willing but the flesh is weak' from English to Russian and back produced 'the vodka is strong but the meat is rotten!' The reason for the failure was that word for word substitution was used. Unfortunately, natural languages are not that simple. More modern approaches use top-down contextual analysis. For example, one program, when given (in America) 'John shot two bucks' will select the meanings 'fired rifle' and 'deer' if primed with 'hunting' but, 'lost' and 'dollars' if primed with 'gambling'.

An associated problem is recognising speech patterns. Modern attempts to interpret spoken English fall into two categories:
1) **template matching** uses the **phonemes** which make up our speech and tries to match them against stored versions;
2) **feature-based** recognition attempts to find speaker independent patterns for linguistic events.

Researchers are also attempting to refine computer vision algorithms. In the previous article we saw how the natural constraints of block world images could be interpreted. Real images are rather more complex. One technique is to differentiate the image in two dimensions, which leaves lines showing the boundaries of objects in the image. These can then be approached in the same way as the block world interpretation.

This technique is inadequate if it is necessary to determine the three dimensional spatial relationships between the objects (as, for example is necessary if a robot is to navigate the streets). If the problem is indoors, then special lighting can simplify the problem. One approach which has been followed is to illuminate the objects with light stripes. The lines so

5



5. A reduction tree architecture evaluates many sub-problems concurrently before passing results up to the next node processor.

formed, model the 3D shape of the objects.

Once the information is installed in the knowledge base, we need to have new techniques for reasoning. Edward Feigenbaum, one of the gurus of AI, has called 'discovery' a myth. It is, he claims,

> 'merely a heuristic search in the space of old concepts and their combinations.'

Whilst not agreeing entirely with this, an early experimental program has invented some mathematics which mathematicians find interesting. (The particular case is that where the program stumbled across the idea of maximally divisible numbers, even though the program's author and other colleagues had never thought about such things. The distinguished mathematician Ramanujan had, however, so the program was in good company.)

Also related to the reasoning mechanism is the problem of reasoning by analogy. This type of reasoning is very widely used by humans. For example, teachers of management give case studies of the type of situations which might arise, and it is then up to the student to recognise the similarities between real life and the example and to extrapolate from the 'solution' appropriately. Most of 'common sense' is based upon this type of reasoning.

Fifth generation hardware

We will now look at the expected form of the new machines that are being developed. The basic hardware of fifth generation machines will probably be an **array** of serial machines linked together in a parallel architecture. Communication between the different processors will be a major factor determining the overall performance of the system.

The type of architecture selected and the scheme used for task allocation are very highly interdependent. It has been said that it might be easier to pick an application, choose how to subdivide the task, and *then* design the machine that it will run on! How can an arbitrary task be subdivided and the parts assigned to individual processors, spending minimal time in performing the subdivision, informing the individual processors, and collecting results? You can see that so much time might be spent in these **overhead** tasks that the performance could be seriously degraded.

Experiments with physical topologies of processors have been taking place for some years. Machines have been devised which use tree structures for inter-processor communications; chessboard patterns with communications links between adjacent processors; rings of processors which pass information on a very high bandwidth bus; multidimensional cubes linked by crossbar networks, and many others.

We shall just consider two strategies which could be followed if processors were arranged in a tree structure. Note, however, that it is not clear just what is the best form of structure to support logical inferences; the examples below is arithemetic.

1) **Reduction** or **demand-driven** machine. *Figure 5* shows a tree of processors, which has, stored at each terminal cell, one element of a large arithmetic problem. Each node processor can perform its part of the calculation and then pass the result up to the next level. All the evaluations at the same level take place simultaneously. The allocation problem is to present the data for the problem to the terminal cells in such a way that the processors have access to the correct information.

2) **Control-driven** machines possess a single centralised source of instructions which are passed down to many processors dealing with parallel streams of data. An example of this approach is the 'Non-Von' machine which will eventually contain 1 million processors connected in a tree. Each tiny processor will have a 1-bit ALU and 64 bytes of memory. Communication is managed via intelligent disks – each processor has a read/write head assigned to it – which are connected to central memory. The central processor sends out instructions to all processors simultaneously.

Conclusion

In these two articles we have surveyed the rapidly moving field of Artificial Intelligence. The coverage has not been exhaustive, but it is hoped that the flavour which has been given will allow you to answer the questions which we posed at the beginning: 'will machines take over the world' and 'will they keep us as pets'.

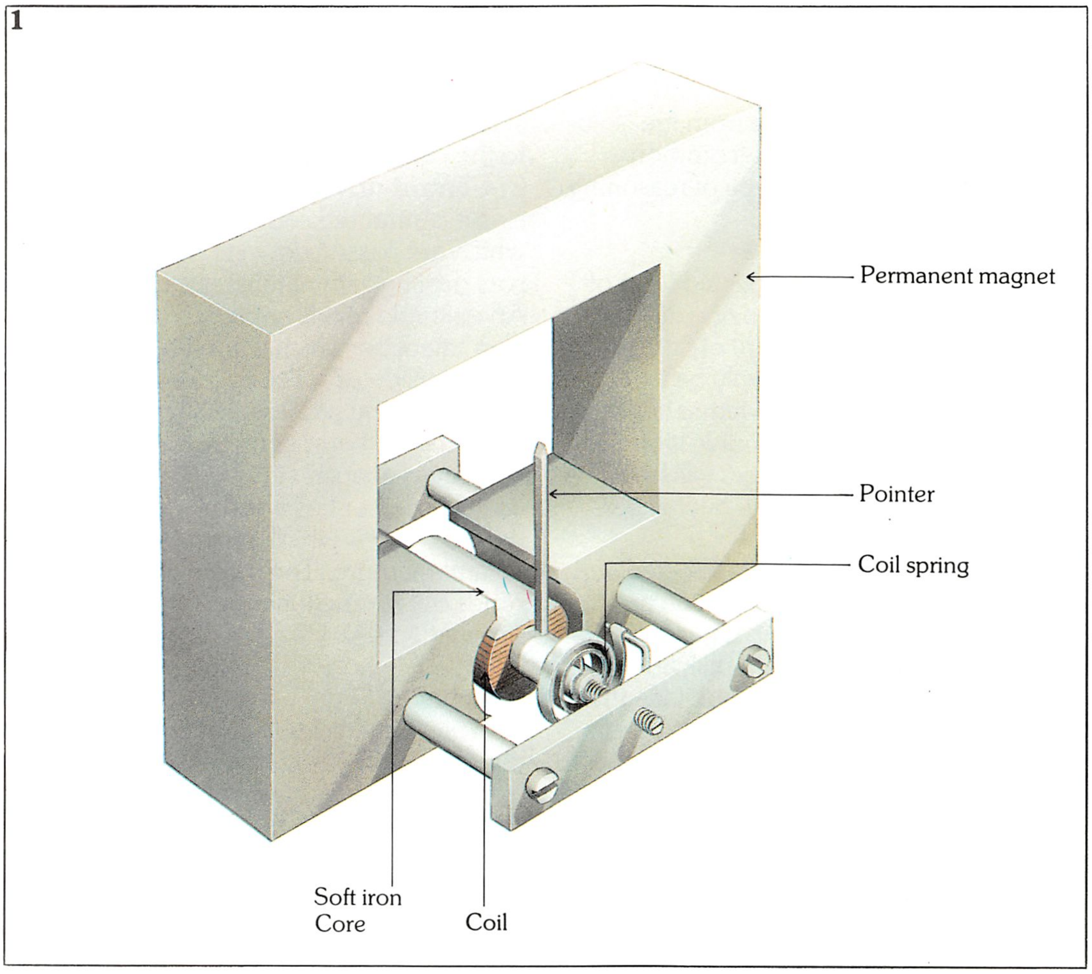
Analogue meters

Analogue meters

Until a few years ago, most measurements in electronics, and the associated field of electrical technology, were undertaken with traditional **moving-coil** instruments; the principle of such a moving-coil meter is shown in *figure 1*. A rectangular coil, wound onto a soft iron core, is mounted on a shaft which is free to rotate between the poles of a permanent magnet. *Figure 2* shows that the circular shape of the magnet's poles produces a radial magnetic field between the poles and the soft iron core. As the coil rotates, therefore, the magnetic flux is always at right angles to the plane of the coil.

Attached to the coil/shaft combination is a pointer which moves as the coil rotates, and a coil spring which ensures that the mechanism always returns to the same point.

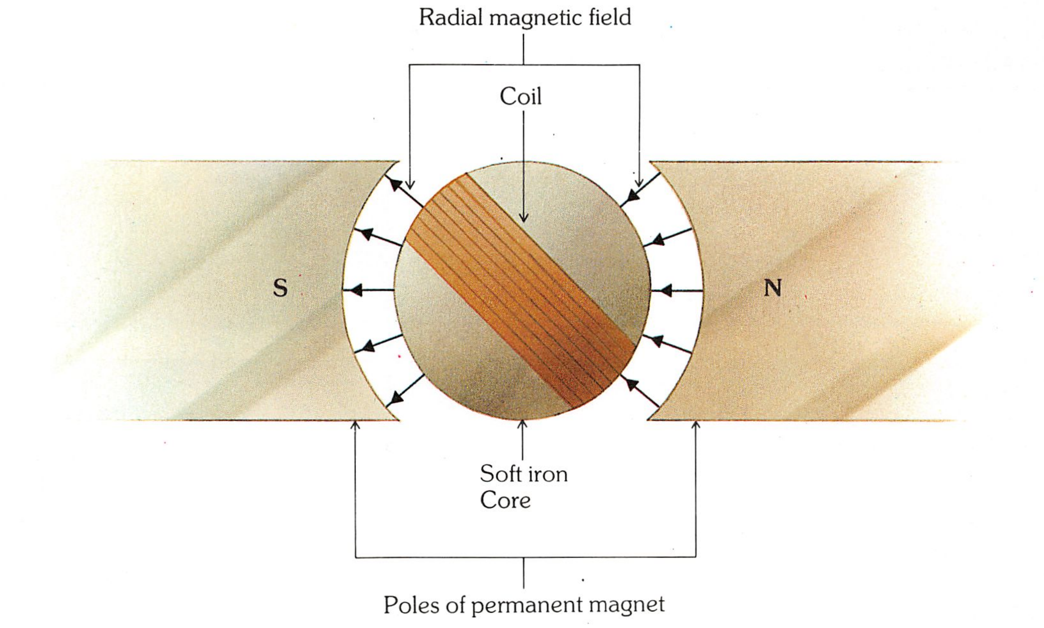
If current passes through the coil, a magnetic field is produced which interacts with the permanent magnetic field, tending to turn the coil/shaft combination, in the same way that the shaft of a motor rotates. If the current through the coil is direct and constant, the coil rotates until the force tending to turn it equals the force of the



1. Principle behind a moving-coil meter.

2. The circular shape of the magnet's poles produces a radial magnetic field between the poles and the soft iron core.

2



spring tending to push it back, at which point the coil is stationary. A larger current thus turns the coil further; a smaller current turns the coil less. In fact, the amount which the coil does turn is linearly proportional to the applied current and so is a direct indication of the magnitude of the current flowing through it.

Associated measurements

We have seen that a moving-coil meter may be used to measure electrical current. It may equally be used to measure the other main electrical quality – voltage. We can see this to be true if we consider Ohm's law, which is summarised by:

$$V = \frac{I}{R}$$

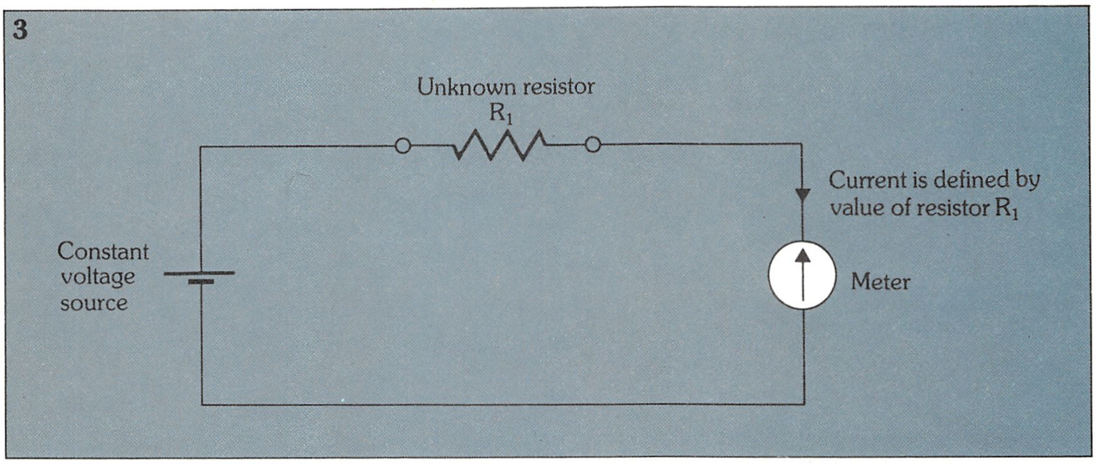
As the main electrical part of the moving coil-meter is the coil itself – which has a resistance – then any current through the coil must produce a voltage across it. Also, any voltage placed across the coil *must produce a current through it.* Voltage across the coil, therefore, causes the coil to rotate, and as the current is directly proportional to the voltage the amount the coil turns is also a direct indication of the magnitude of the voltage.

The moving-coil meter may also be used to indicate resistance. *Figure 3* illustrates a basic circuit in which a meter is used with a voltage source to indicate the value of resistor R_1. For example, if we say that the voltage source provides a voltage of V volts, the current through the coil is therefore (from Ohm's law):

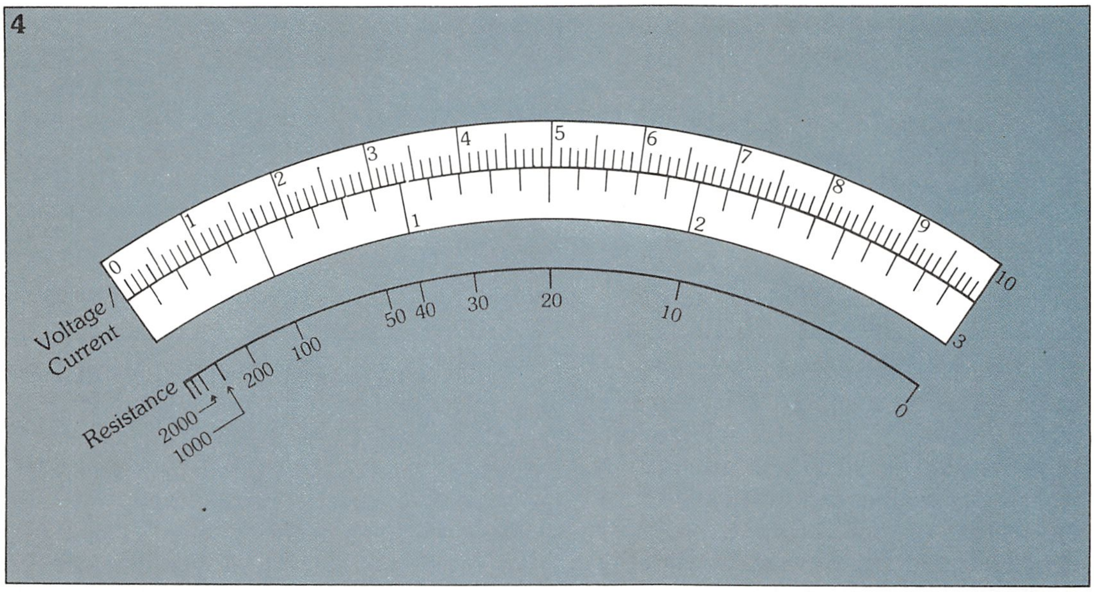
$$I = \frac{V}{R}$$

where R is, in fact, the sum of resistor R_1 and the resistance of the meter coil. This simple relationship illustrates one very important point about the use of a moving-coil meter to indicate resistance value: as the movement of the coil is proportional to the current through it, and as the current is inversely proportional to the resistance, then the movement of the coil is also *inversely proportional to the resistance.* That is, the coil will move *most* for resistances of low values, and *least* for resistances of high values.

This inverse relationship between resistance and movement of the coil means that the movement is non-linear, i.e. equal resistance changes do not produce equal coil movements. An example of a meter



3. A basic circuit which indicates how a moving-coil meter can be used to measure resistance.



4. A non-linear resistance scale and two linear voltage/current scales.

scale which is marked to indicate resistance is shown in *figure 4*. We can see that the further round the scale the pointer moves in a clockwise direction, the lower the resistance indication. Non-linearity of the scale is also evident.

Figure 4 also shows two possible scales for voltage and current indication. These *are* linear and increase as the pointer moves clockwise.

Other measurements

A meter which is capable of indicating voltage, current and resistance may easily be used to indicate a non-electronic quantity as long as a transducer is available to convert a measure of the quantity into a quantity the meter *can* indicate.

Temperature, for example, may be indicated using a moving-coil meter, if a thermocouple is available to convert a measure of the temperature to voltage. We have seen such a thermocouple in our previous study of instrumentation systems.

Oil pressure transducers which convert a measure of engine oil pressure to resistance – an electronic quantity which the moving-coil meter can indicate – form another example.

Using one meter for many measurements

The coil in a moving-coil meter may only rotate through a certain angle. Generally, a mechanical stop prevents rotation through an angle of more than, say, 90°. The current required through the coil to produce this maximum (known as **full scale deflection**, FSD) is known as the meter's **sensitivity**. Sensitivities as low as 50 μA FSD are common.

As the coil of a meter has resistance,

however, the meter's sensitivity may also be expressed as the voltage required to produce full scale deflection. A meter with a current sensitivity of, say, 1 mA, and a coil resistance of 125 Ω, would have a voltage sensitivity of:

$$\begin{aligned} V &= 1 \times 10^{-3} \times 125 \\ &= 125 \text{ mV FSD} \end{aligned}$$

Although it may sometimes be necessary to measure currents and voltages as small as this (and occasionally *smaller* than this) it is more usual that the currents and voltages in typical electronic circuits will be much larger. However, measuring large currents and voltages, directly, with a moving-coil meter with a sensitivity like this, would result in too much current through the coil, dissipating excessive power which would melt the wire forming the coil. That is, the coil would burn out and be destroyed.

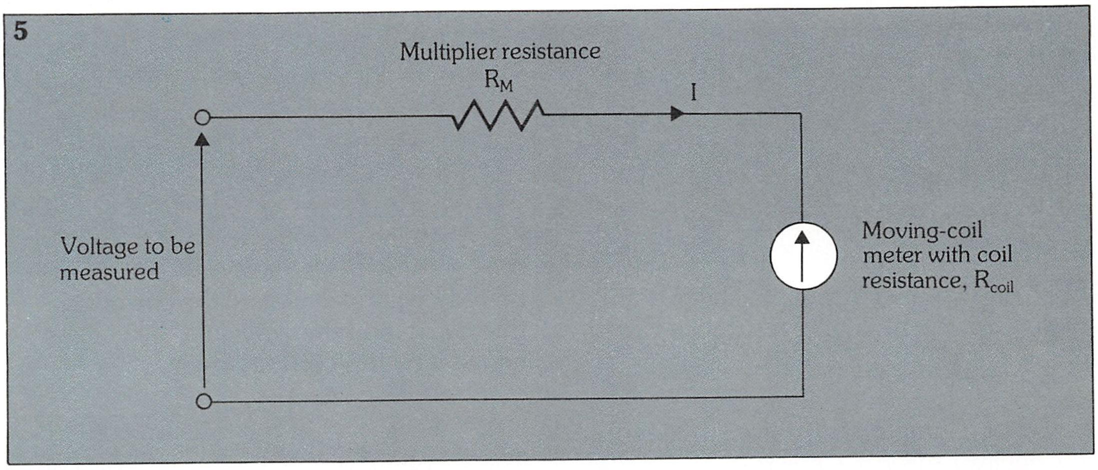
We could make several moving-coil meters, each with a different sensitivity, and use whichever sensitivity is required for a particular current or voltage measurement – but this would be very expensive. For a range of voltages of, say, 250 mV to 100 V, possibly, 5 meters would be needed. A much cheaper and more desirable alternative is to adapt a single highly sensitive meter and provide it with many different **ranges** between the minimum and maximum measurements required.

Extra voltage ranges, for example, are easily obtained by adding resistors in series with the meter coil. These series resistors are known as **multipliers**. *Figure 5* shows a meter which has been adapted to a higher voltage range with a multiplier resistance.

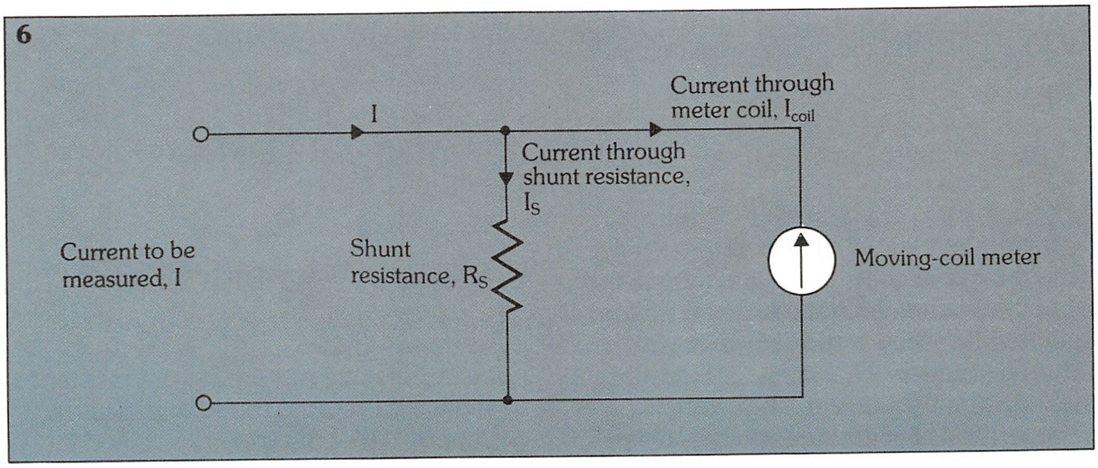
The value of the multiplier depends on the voltage range to be measured, and, of course, on the meter's coil resistance. It is a simple matter to calculate the required multiplier. From Ohm's law:

$$R = \frac{V}{I}$$

where R is the combined resistance of the multiplier and the coil, V is the required

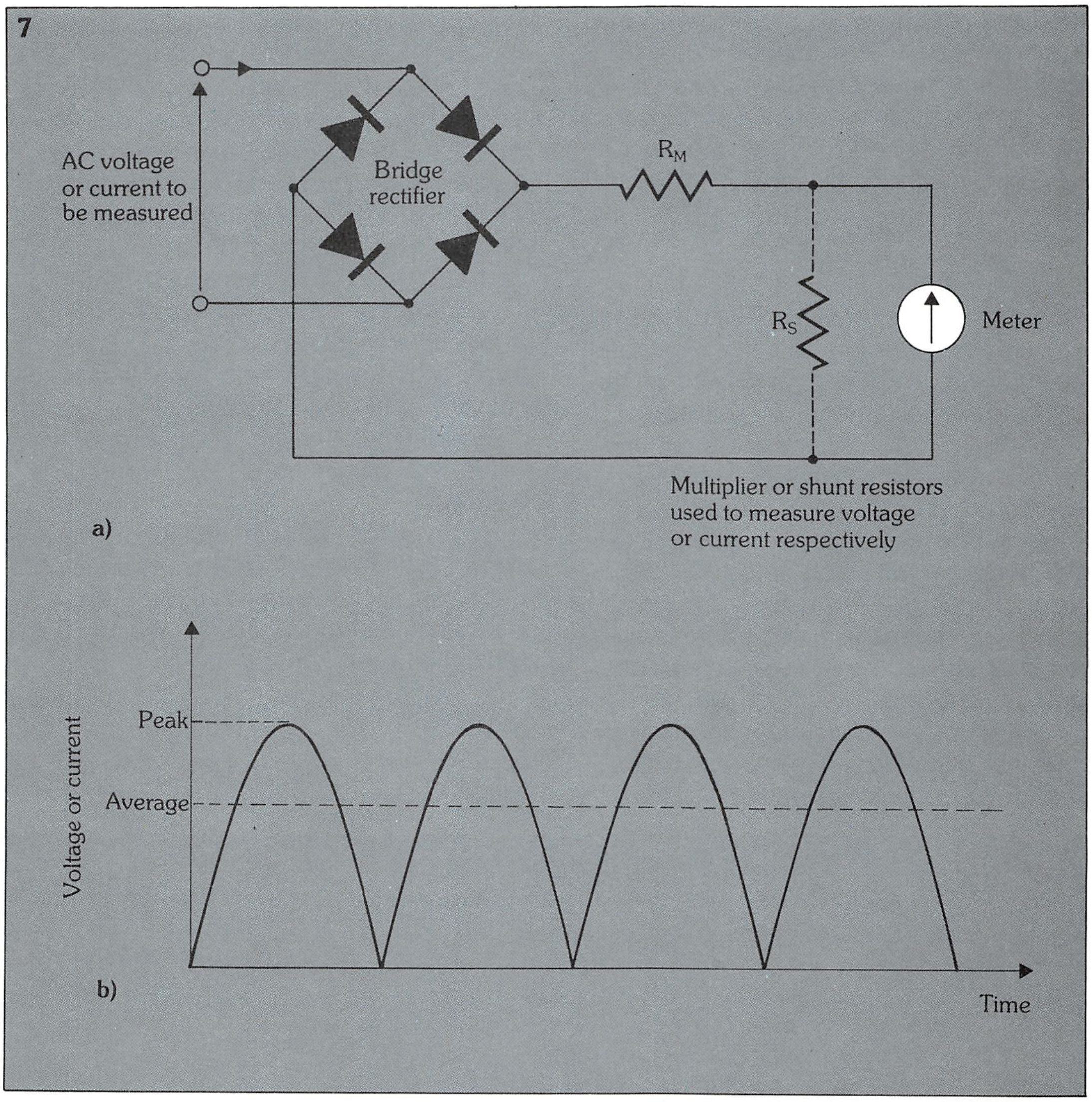


5. Adding a multiplier resistance to a moving-coil meter enables it to measure higher voltage ranges.



6. Shunt resistors (resistors in parallel) enable a moving-coil meter to measure higher current values.

7



7. (a) Using a full-wave bridge rectifier to rectify the alternating current to DC; **(b)** graph of voltage or current against time showing the average of the alternating current is 0.636 of the peak magnitude.

FSD, and I is the current sensitivity of the meter. So, the multiplier resistance R_M, is:

$$R_M = \frac{V}{I} - R_{coil}$$

For example, if we have a meter of sensitivity 1 mA FSD, coil resistance 125 Ω and we wish to use it to indicate a voltage of 15 V FSD, the multiplier resistance is:

$$R_M = \frac{15}{1 \times 10^{-3}} - 125$$
$$= 1.5 \times 10^4 - 125$$
$$= 14.875\ \text{k}\Omega$$

Similarly, extra current ranges may be obtained by adding resistors *in parallel* with the meter coil – these are known as **shunt resistors**. *Figure 6* shows a meter adapted to a higher current range with a shunt resistance. The value of the shunt depends on the current range to be measured. The shunt resistance is determined by first calculating the current I_S, through it, and then dividing the voltage across it by this current. Thus:

$$R_S = \frac{V_S}{I_S}$$

The current through the shunt is, from Kirchhoff's law:

$$I_S = I - I_{coil}$$

where I is the required current sensitivity.

The voltage across the shunt resistor is the same as the voltage across the coil producing FSD, so:

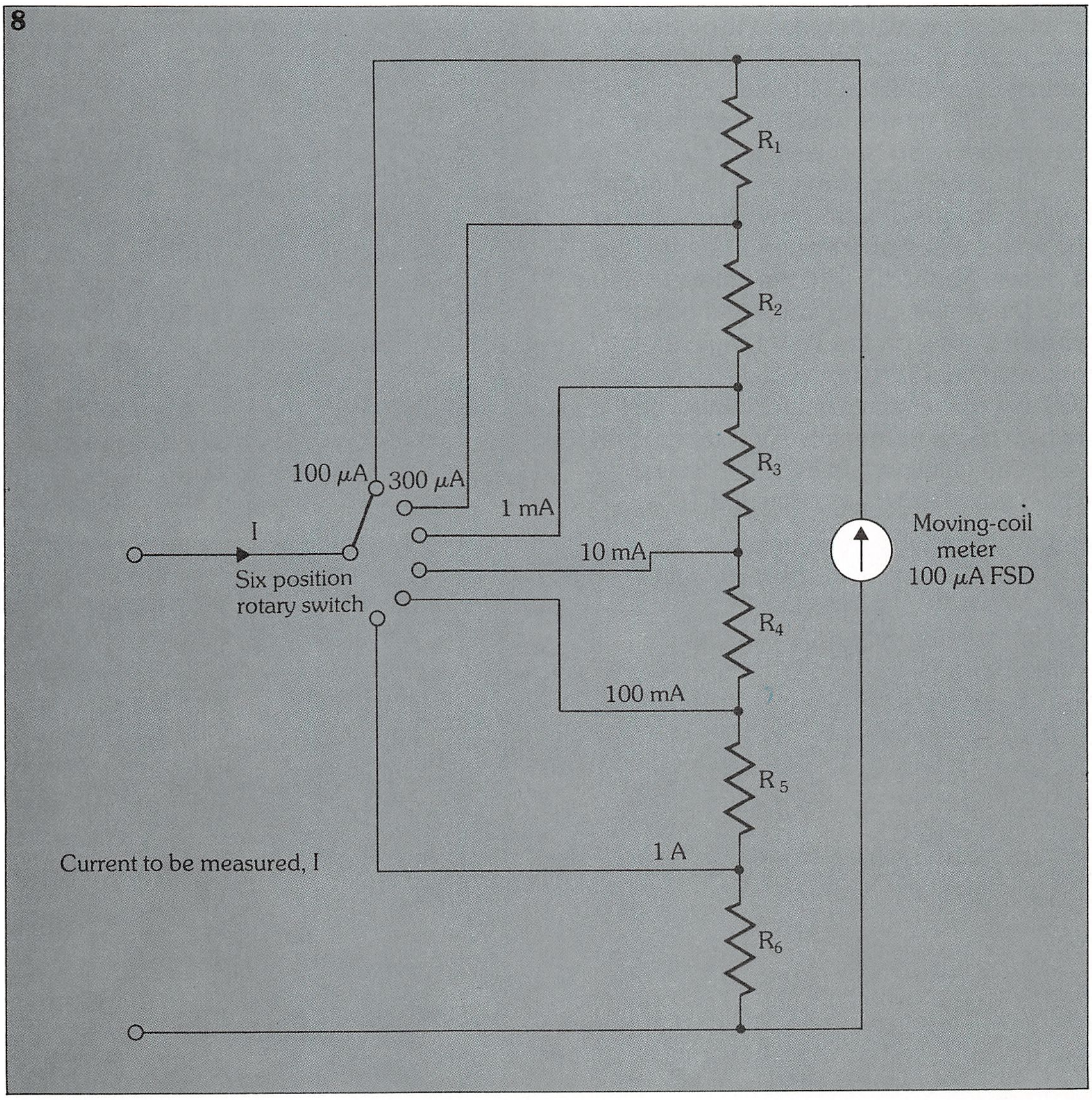
$$V_S = V_{coil}$$

and, therefore, the shunt resistance is:

$$R_S = \frac{V_{coil}}{I - I_{coil}}$$

For example, the required shunt resistance to allow the previous meter to indicate a current of 10 A FSD is:

8



8. Possible circuit of a multirange ammeter.

$$R_S = \frac{0.25}{10 - 1 \times 10^{-3}}$$
$$= 0.025025\ \Omega$$

A resistor of 0.025 Ω will ensure that the error is not greater than 0.1%.

AC measurements

The moving-coil meter we have described is only capable of indicating direct currents and voltages. An alternating current voltage will attempt to make the coil rotate first in one direction, then in the other, but as a consequence of the inertia of the instrument the result is that the meter indicates the *average value* – which for a sine wave, for example, is zero.

A **full-wave bridge rectifier** may be used to rectify the alternating current to be measured to DC, as shown in *figure 7a*, thus the average of the *magnitude* of the alternating current is indicated. The average of the AC magnitude is 0.636 of the peak magnitude (*figure 7b*). Normally, however, a meter would be required to indicate the rms value which is 0.707 of the peak value, so the meter scale should be marked to read:

$$\frac{0.707}{0.636} = 1.11$$

times the average indication, or the accompanying values of shunts and multipliers must be adjusted to suit.

Multirange meters

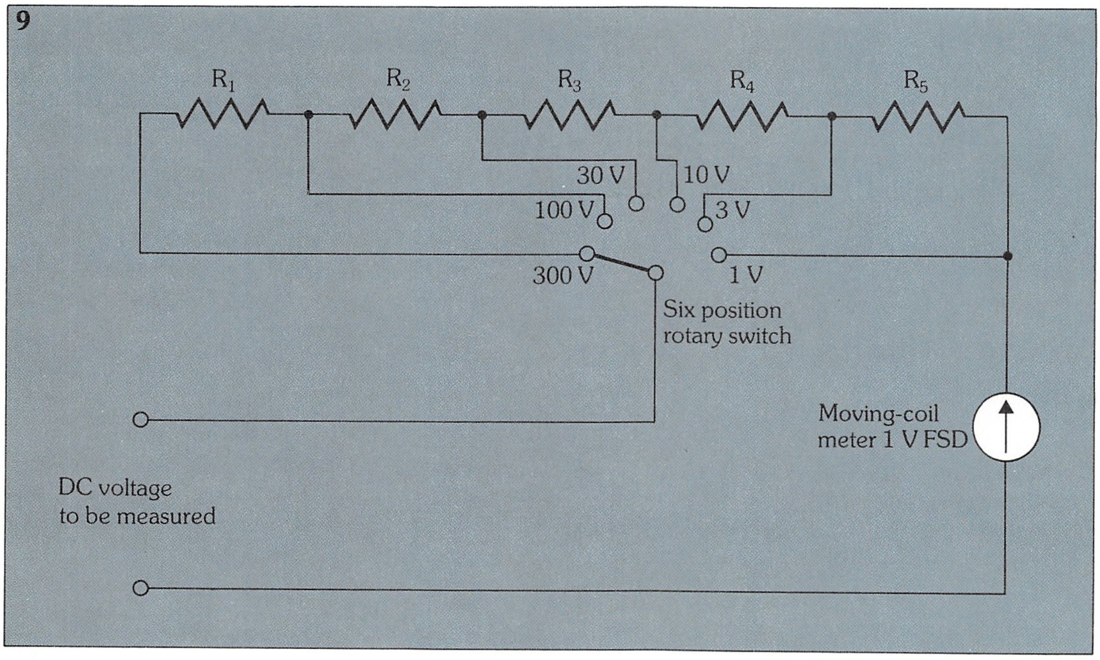
Certain types of meters exist (based on the principles we have seen here of moving-coil meters with shunts, multipliers and voltage sources) which allow a user to measure a variety of voltages, currents and resistances simply by selecting the range

required, on a switch mechanism. Such meters are often called **multimeters**, a name which implies their multipurpose role. Typical ranges which a multimeter may measure are shown in *table 1*.

The accuracy of a typical multimeter, i.e. the closeness of the indicated value to the actual value of the measured quantity, is usually about ±1% of the full scale value over the whole range. This means that when it is, say, on the 10 V range, the indicated value is within 1% of 10 V, i.e. 100 mV of the actual value. However, if a voltage measurement of 1 V is taken, the indicated value would be 1 V ± 100 mV – which is a reading accuracy of ±10%.

Table 1
Range of multimeter measurements

Voltage	Current	Resistance
DC: 300 mV	DC: 100 μA	1 kΩ
1 V	300 μA	100 kΩ
3 V	1 mA	1 MΩ
10 V	3 mA	
30 V	10 mA	
100 V	100 mA	
AC: 300 mV	1 A	
1 V	AC: 100 mA	
3 V	1 A	
10 V		
30 V		
100 V		



9. A multirange voltmeter.

Obviously, then, a multimeter is most accurate when the measured value is close to the FSD of the range: an important point when using a multimeter.

A possible circuit of a multirange ammeter is shown in *figure 8*, where a number of shunts are switched into and out of circuit to give the required range FSD. A multirange voltmeter is shown in *figure 9*. Here, multipliers are switched into and out of circuit.

A multimeter is constructed using circuits similar to those of *figures 8* and *9*, along with a rectifier (for AC measurements) and a battery (the voltage source for resistance measurements) in a single case. Many such multimeters are available, varying considerably in price, according to facilities, accuracy, sensitivity and overall quality and robustness.

Because of its range of uses and functions, the multimeter is the most important member of a family of measuring equipment known collectively as **test equipment**. We shall consider other items of test equipment in later *Digital and Solid State Appendices*.

Electronic meters

The moving-coil meters we have seen so far all take the power required to rotate the meter coil from the circuit whose voltage or current is being measured. In most circuits this presents little or no problem because

the meter power current is only a small proportion of the total. However, in some circuits, the meter may load the measured circuit to such an extent that the indicated voltage is highly inaccurate. The circuit in *figure 10* is an example.

According to the potential divider rule, the voltage across resistor R_1, when the meter is not connected, is given by:

$$V_{R1} = \frac{R_1}{R_1 + R_2} V_{in}$$

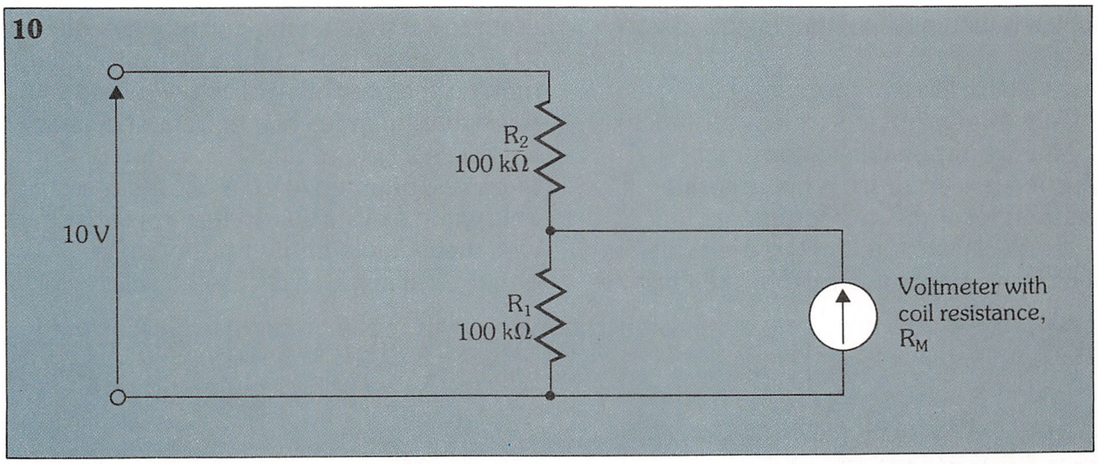
$$= \frac{100{,}000}{100{,}000 + 100{,}000} \times 10$$
$$= 5\,V$$

When the meter is connected, however, its resistance is placed in parallel with resistor R_1 producing an effective resistance of:

$$R_{eff} = \frac{R_1 R_M}{R_1 + R_M}$$

where R_M is the meter resistance. If, for example, the meter resistance is10 kΩ, then the effective resistance of the two

10. In this circuit the meter loads the measured circuit to such an extent that the meter reading is highly inaccurate.



Below: an analogue multimeter.

The Research House/Thorn-EMI Instruments Ltd

parallel resistors is therefore:

$$R_{eff} = \frac{100{,}000 \times 10{,}000}{100{,}000 + 10{,}000}$$
$$= 9.091\,k\Omega$$

The voltage indicated by the meter is now given by the potential divider rule as:

$$V_{R1} = \frac{9091}{100{,}000 + 9091} \times 10$$
$$= 0.83\,V$$

which is much lower than the correct voltage when the circuit is not loaded by the meter.

This leads to an imporant question: what circuits may we use a moving coil meter (of the type we have seen here) with?

Well, the answer is not simple, but a general rule-of-thumb is that the resistance across which the voltage is to be measured, must be, at most, 0.1. of the meter resistance. Or, put another way, the meter resistance must be *at least* 10 times that of the circuit resistor. To take an example, let's say the meter resistance R_M of *figure 10* is 1 MΩ (i.e. 10 times resistor R_1), then the effective resistance, R_{eff}, of resistors R_1 and R_M in parallel is:

$$R_{eff} = \frac{R_1 R_M}{R_1 + R_M}$$

$$= \frac{100{,}000 \times 1{,}000{,}000}{100{,}000 \times 1{,}000{,}000}$$

$$= 90.91 \text{ k}\Omega$$

and the measured voltage is:

$$V_{R1} = \frac{90910}{100{,}000 + 90910} \times 10$$

$$= 4.76 \text{ V}$$

which is within 5% of the unloaded voltage.

There are two problems here:
1) moving-coil meters with a resistance of 1 MΩ are impossible to construct;
2) even 5% accuracy is not accurate enough for some measurements.

Electronic multimeters are available which overcome these problems, by interfacing the moving-coil meter movement to the circuit being tested with an electronic circuit. The electronic circuit is designed with an extremely high input resistance (often in the order of 10 MΩ and above) which ensures that the tested circuit is not loaded and that accuracy is not affected.

Interfacing circuits of such electronic voltmeters are typically based upon transistor operation (generally FET) or op-amps. Accuracy of such electronic analogue multimeters is however, still limited by the accuracy of the moving-coil movement.

An advantage of the electronic multimeter is the possibility of its use measuring AC signals of much higher frequency than a non-electronic multimeter. Input frequencies, often up to 10 MHz, are measurable with an electronic meter, compared with measurable frequencies of up to about 10 kHz with a non-electronic meter.

Glossary

full scale deflection	maximum value, indicated by a measuring instrument, of a measured quantity
moving-coil meter	a meter movement, consisting of a coil suspended in a permanent magnetic field. Current through the coil causes it to rotate through an angle, dependent on the magnitude of the current
multimeter	measuring instrument capable of indicating a number of ranges of voltages, currents and resistances
multiplier	a series resistor which allows a moving-coil meter to be used to indicate higher voltage ranges
ranges	values indicated by full scale deflection of a multimeter
sensitivity	the current through the coil of a moving-coil meter which produces full scale deflection of the coil
shunt	resistor in parallel with a meter movement which allows the movement to indicate higher ranges of current

ELECTRICAL TECHNOLOGY

Power in a three-phase system

The power in a three-phase system can be determined in a similar way to that used for single-phase systems. The total instantaneous power in a three-phase load, p, may be written:

$$P = v_R i_R + v_Y i_Y + v_B i_B$$

Where v_R, i_R, v_Y, i_Y, v_B and i_B are the instantaneous voltages and currents of the three phases of a load which we can assume is connected in a star.

The average power, P, can be written as the sum of the average powers in the three phases:

$$P = V_R I_R \cos\phi_R + V_Y I_Y \cos\phi_Y + V_B I_B \cos\phi_B$$

Where $V_R I_R \cos\phi_R$ etc. are the rms voltage, current and power factor in each of the three phases of the star connected load impedances, as shown in *figure 1*.

The reactive power, Q, can be determined as:

$$Q = V_R I_R \sin\phi_R + V_Y I_Y \sin\phi_Y + V_B I_B \sin\phi_B$$

and we must remember that some of these terms may be positive and some negative, depending on the values of the phase angles.

The apparent power, S, can be found from the expression:

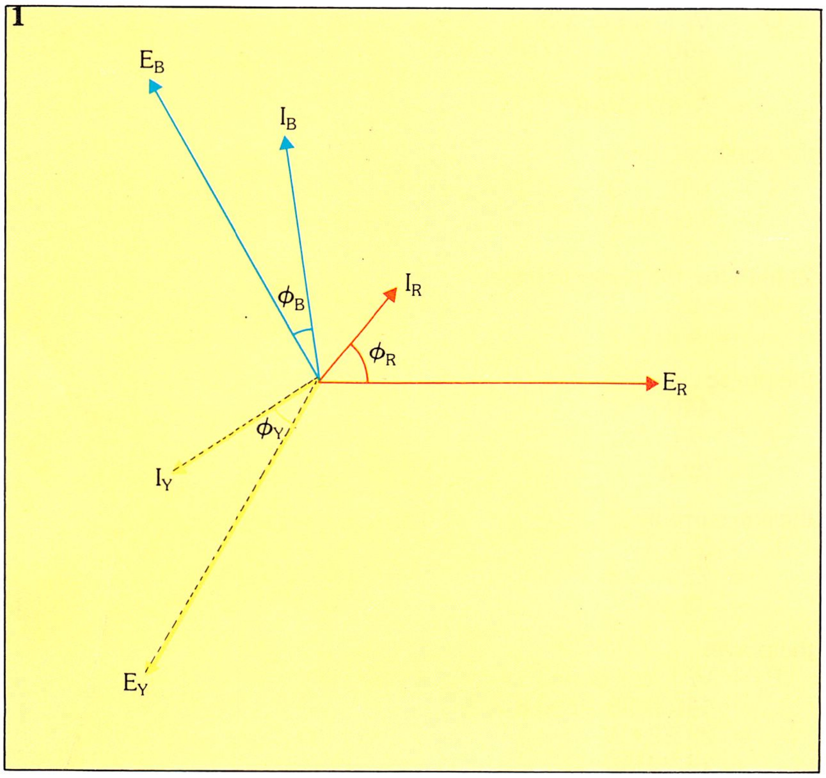
$$S = \sqrt{P^2 + Q^2}$$

(In a three-phase system, the apparent power cannot be determined from the sum of the apparent power in the three separate phases.)

The power factor for the complete three-phase load can be determined as:

$$\cos\phi = \frac{P}{S}$$

1. Phasor diagram showing the rms voltage, current and power factor in each of the 3 phases of the star connected load.



Power in a symmetrically balanced star

In a symmetrical system:

$E_R = E_Y = E_B = E_P$ (the phase voltage)

again for a balanced load:

$$I_R = I_Y = I_B = I_P$$

and the phase angle of each branch is also the same:

$$\phi_R = \phi_Y = \phi_B = \phi_P$$

Where ϕ_P is the phase angle of each branch of the load. The average power, P, can be obtained from the same expression used for an unbalanced system:

$$P = 3V_P I_P \cos\phi_P$$

Remember, the phase voltage V_P, is related to the line voltage, V_L by:

$$V_L = V_P\sqrt{3}$$

and that $V_L = I_L$ for a star connected load. This gives us:

$$P = V_L I_L \cos\phi_P \sqrt{3}$$

and similarly:

$$Q = 3V_P I_P \sin\phi_P$$

$$= V_L I_L \sin\phi_P \sqrt{3}$$

We can obtain the apparent power, S, from this as:

$$S = \sqrt{P^2 + Q^2}$$

$$= \sqrt{3V_L^2 I_L^2 \cos^2\phi + 3V_L^2 I_L^2 \sin^2\phi}$$

$$= V_L I_L \sqrt{3}$$

since $\cos^2\phi + \sin^2\phi = 1$ for all values of ϕ. This balanced load's power factor is given by:

$$\cos\phi = \frac{P}{S}$$

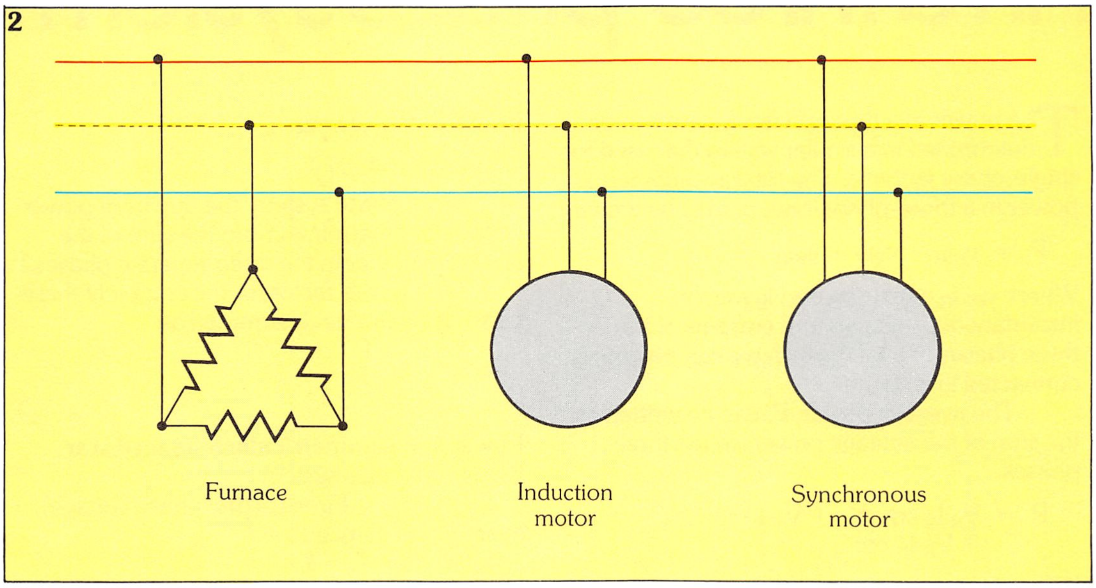
$$= \frac{V_L I_L \cos\phi_P \sqrt{3}}{V_L I_L \sqrt{3}}$$

$$= \cos\phi_P$$

So we can see that the power factor for the complete balanced load is identical to the power factor of each phase of the load.

Power in a balanced delta

In a delta connected load, the average power is three times that in each branch, thus:



2. A furnace, induction motor and synchnonous motor connected in parallel to a three-phase supply.

$$P = 3V_P I_P \cos \phi_P$$
and:
$$Q = 3V_P I_P \sin \phi_P$$
V_P, I_P and ϕ_P are the magnitudes of voltage and current and the phase angle for the three separate branches of the delta connected load.
Since in this case:
$$V_L = V_P$$
and:
$$I_L = I_P \sqrt{3}$$
we obtain:
$$P = V_L I_L \cos \phi_P \sqrt{3}$$
and:
$$Q = V_L I_L \sin \phi_P \sqrt{3}$$
From these we can see that:
$$S = V_L I_L \sqrt{3}$$
and that again $\cos \phi = \cos \phi_P$

So we can see that it doesn't matter how the load is connected to a three-phase supply, as the real power absorbed by any three-phase load is:
$$P = V_L I_L \cos \phi \sqrt{3}$$
However, this of course doesn't mean that the power absorbed by a set of three impedances is the same, if they are connected in star or delta, as we shall see.

We'll now determine the current, power and reactive power when a three-phase load consisting of three impedances of 20 Ω and power factor of 0.8 is connected in star and delta to a three-phase 440 V supply:

1) In star, the phase voltage:
$$V_P = \frac{440}{\sqrt{3}} = 254\text{ V}$$
the line current:
$$I_L = I_P = \frac{254}{20} = 12.7\text{ A}$$
the real power:
$$P = V_L I_L \cos \phi \sqrt{3} = 440 \times 12.7 \times 0.8 \times \sqrt{3} = 7743\text{ W} = 7.74\text{ kW}$$
the reactive power:
$$Q = V_L I_L \sin \phi \sqrt{3} = 440 \times 12.7 \times 0.6 \times \sqrt{3} = 5807\text{ VAR} = 5.807\text{ kVAR}$$
the apparent power:
$$S = \sqrt{P^2 + Q^2} = 9.68\text{ kVA}$$

2) In delta, the phase voltage:
$$V_P = V_L = 440\text{ V}$$
the phase current:
$$I_P = \frac{440}{20} = 22\text{ A}$$
the line current:
$$I_L = I_P \sqrt{3} = 22 \times \sqrt{3} = 38.1\text{ A}$$
the power:
$$P = V_L I_L \cos \phi \sqrt{3} = 440 \times 38.1 \times 0.8 \times \sqrt{3} = 23{,}229\text{ W} = 23.23\text{ kW}$$

the reactive power:

$$
\begin{aligned}
Q &= V_L I_L \sin\phi \sqrt{3} \\
&= 440 \times 38.1 \times 0.6 \times \sqrt{3} \\
&= 17{,}421 \text{ VAR} \\
&= 17.42 \text{ kVAR}
\end{aligned}
$$

the apparent power:

$$
\begin{aligned}
S &= \sqrt{P^2 + Q^2} \\
&= 29.04 \text{ kVA}
\end{aligned}
$$

Power in parallel three-phase loads

The power in a group of three-phase loads connected in parallel can be determined in a similar way to that which was used for single-phase systems.

As an example, consider an electric furnace that consumes power of 15 kW at unity power factor; a three-phase induction motor consuming 20 kW at a lagging power factor of 0.75 ; and a synchronous motor consuming power of 5 kW at a leading power factor of 0.6. These are connected in parallel to a three-phase supply as shown in *figure 2*.

The furnace:

$$
\begin{aligned}
P_1 &= 15 \text{ kW} \\
Q_1 &= 0 \text{ VAR}
\end{aligned}
$$

The induction motor.

$$
\begin{aligned}
\cos\phi &= 0.75 \\
\phi &= 41.41^\circ \\
\sin\phi &= 0.66 \\
P_2 &= 20 \text{ kW} \\
S_2 &= \frac{20}{0.75} \\
&= 26.67 \text{ kVA} \\
Q_2 &= 26.67 \times 0.66 \\
&= 17.60 \text{ kVAR}
\end{aligned}
$$

The synchronous motor:

$$
\begin{aligned}
\cos\phi &= 0.6 \\
\phi &= 53.1^\circ \\
\sin\phi &= -0.8 \text{ (since the power factor is leading)} \\
P_3 &= 5 \text{ kW} \\
S_3 &= \frac{5}{0.6} \\
&= 8.33 \text{ kVA} \\
Q_3 &= -6.67 \text{ kVAR}
\end{aligned}
$$

The total lead can be computed from:

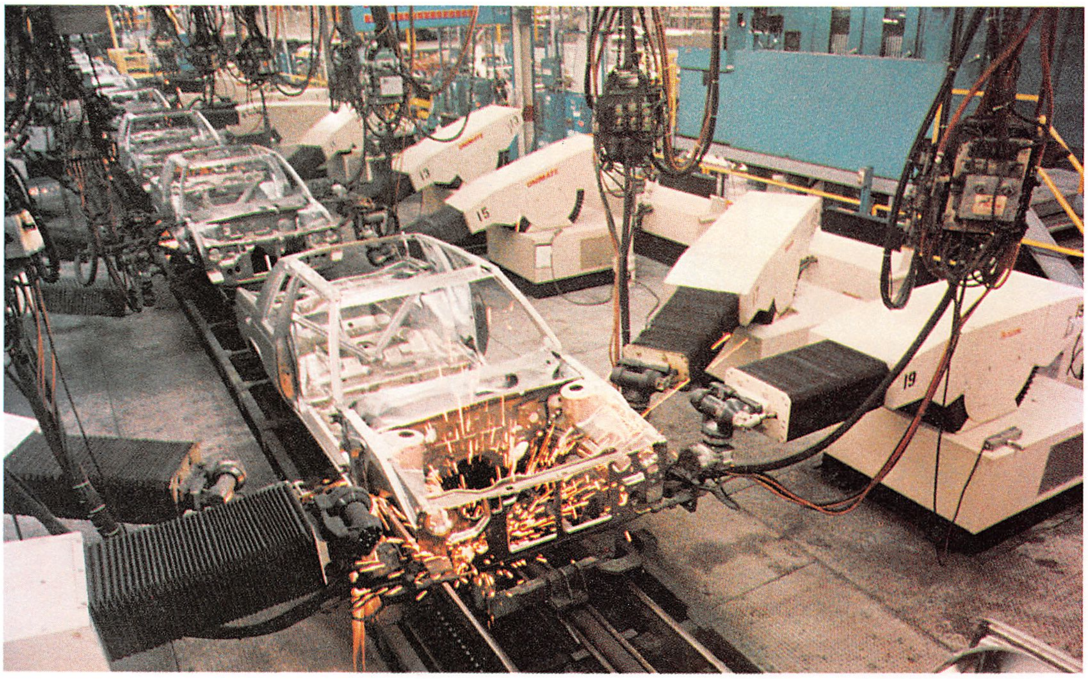
$$
\begin{aligned}
P &= P_1 + P_2 + P_3 \\
&= 15 + 20 + 5 \\
&= 40 \text{ kW}
\end{aligned}
$$

$$
\begin{aligned}
Q &= Q_1 + Q_2 + Q_3 \\
&= 0 + 17.60 - 6.67 \\
&= 10.93 \text{ kVAR}
\end{aligned}
$$

$$
\begin{aligned}
S &= \sqrt{40^2 + 10.93^2} \\
&= 41.47 \text{ kVA}
\end{aligned}
$$

$$
\begin{aligned}
\text{The power factor} &= \frac{P}{S} \\
&= \frac{40}{41.47} \\
&= 0.965
\end{aligned}
$$

□



Right: a robot welding assembly line at the Chrysler works, Missouri.

Science Photo Library/Tom McHugh

4 COMMUNICATIONS

Television and radio

Television and radio – the media

Television has opened our eyes. We take for granted its matter-of-fact coverage of events – from around the world, from space, and from the depths of the ocean. Prior to television, radio was the primary means of mass communication – relying perhaps on the listener's imagination to provide the colour and atmosphere supplied by television's video signal.

Both of these communications media convert vision or sound information into electrical pulses which (in the right form and strength) may be radiated from a broadcasting aerial or sent down a wire or cable.

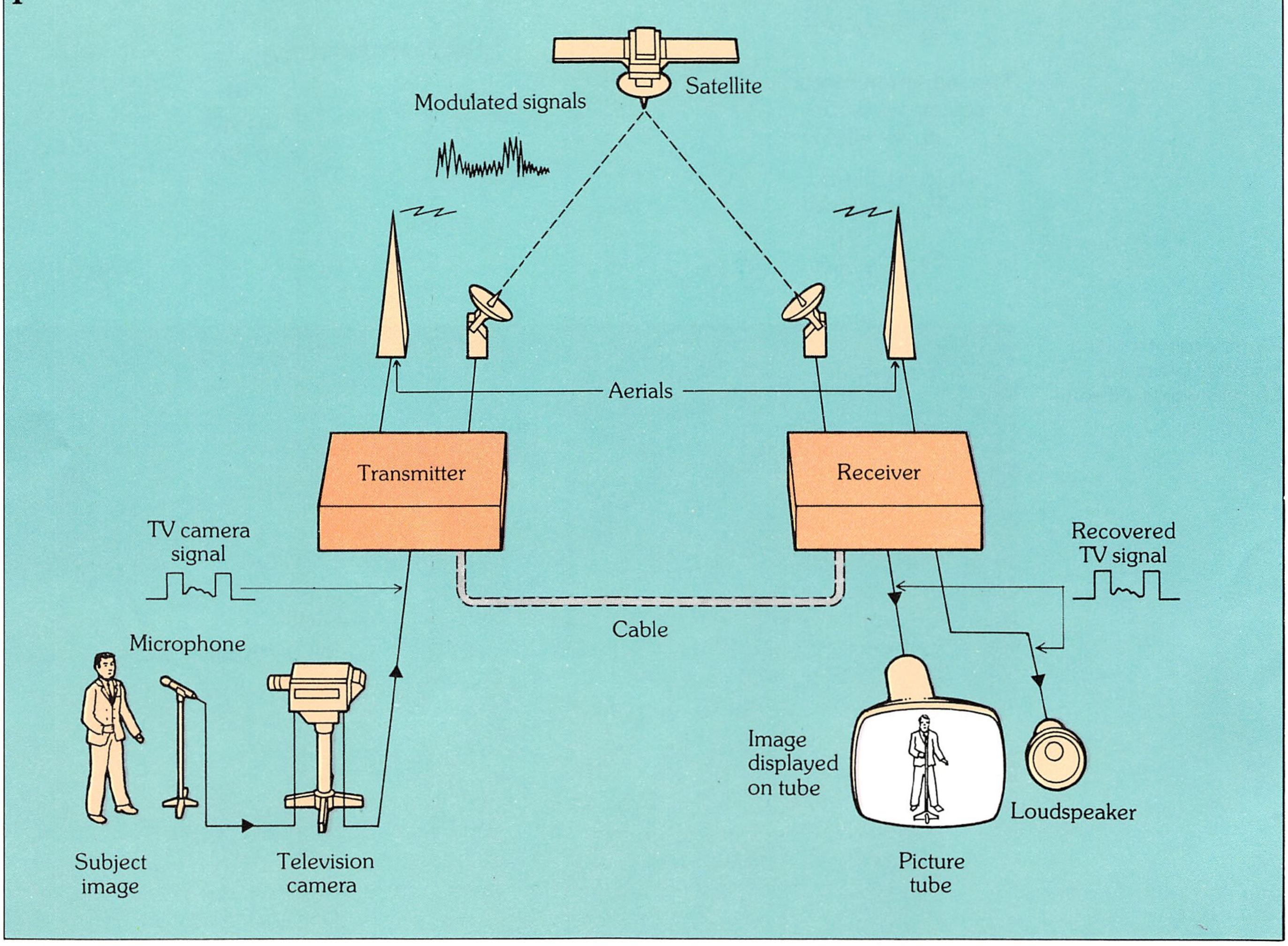
We'll take a detailed look at television first, because many of the principles involved also apply to radio.

Principles of television

Today's television picture is an illusion – in reality, the U.K. viewer looks at fifty 'half

1. Basic components of a television transmission and receiving system.

1



pictures' each second, and unconsciously relies on the eye and brain to fuse them into an apparently steady image. This illusion relies on the combination of a rapidly moving pin-point of bright light and the eye's persistence of vision. The pin-point of light scans the screen so quickly that the eye and brain fuse the millions of separate pin-point positions into one image.

The basic building blocks of a TV transmission and receiving system are shown in *figure 1*. The television camera converts the optical scene – more correctly the reflected light waves from the scene – into electrical signals, and the microphone alongside converts the audio scene into corresponding electrical signals. These signals are modulated onto a high frequency carrier signal, along with synchronising information, and the combined signal is amplified so that it may radiate electromagnetically from a broadcasting aerial – or bounce off a satellite, or travel along a cable.

At the receiver, the signal is demodulated, amplified, and the video and sound signals processed to produce the picture and sound that the viewer has tuned in to see.

In order to understand how a television picture is transmitted and displayed, we'll look first at black and white television. Suppose we wanted to transmit the image of a chessboard, electrically, from one room to another. One method of doing this would be to arrange 64 light bulbs in an eight by eight grid to represent the chessboard squares to be transmitted, and connect each bulb and its battery through a switch to a matching set of bulbs in the other room. If we constructed a code which said: 'if it's lit up, then it's a white square', it would be easy to duplicate the board's alternating pattern of black and white squares. Indeed any pattern could be displayed; even letter shapes may be represented and messages slowly exchanged.

To take this simple arrangement one stage further, we could replace each transmitting bulb with a light sensitive cell and connect the circuit so that a lamp only illuminates if light is falling on the face of its corresponding cell. The pattern of the chessboard would then be projected through a lens onto the face of the cells and the resulting electrical signals would then be amplified, transmitted, and duplicated on the bulbs. The greater the number of cells and bulbs, the finer the detail of the picture that could be transmitted.

However, the complications of wiring and low currents would make such a scheme unrealistic.

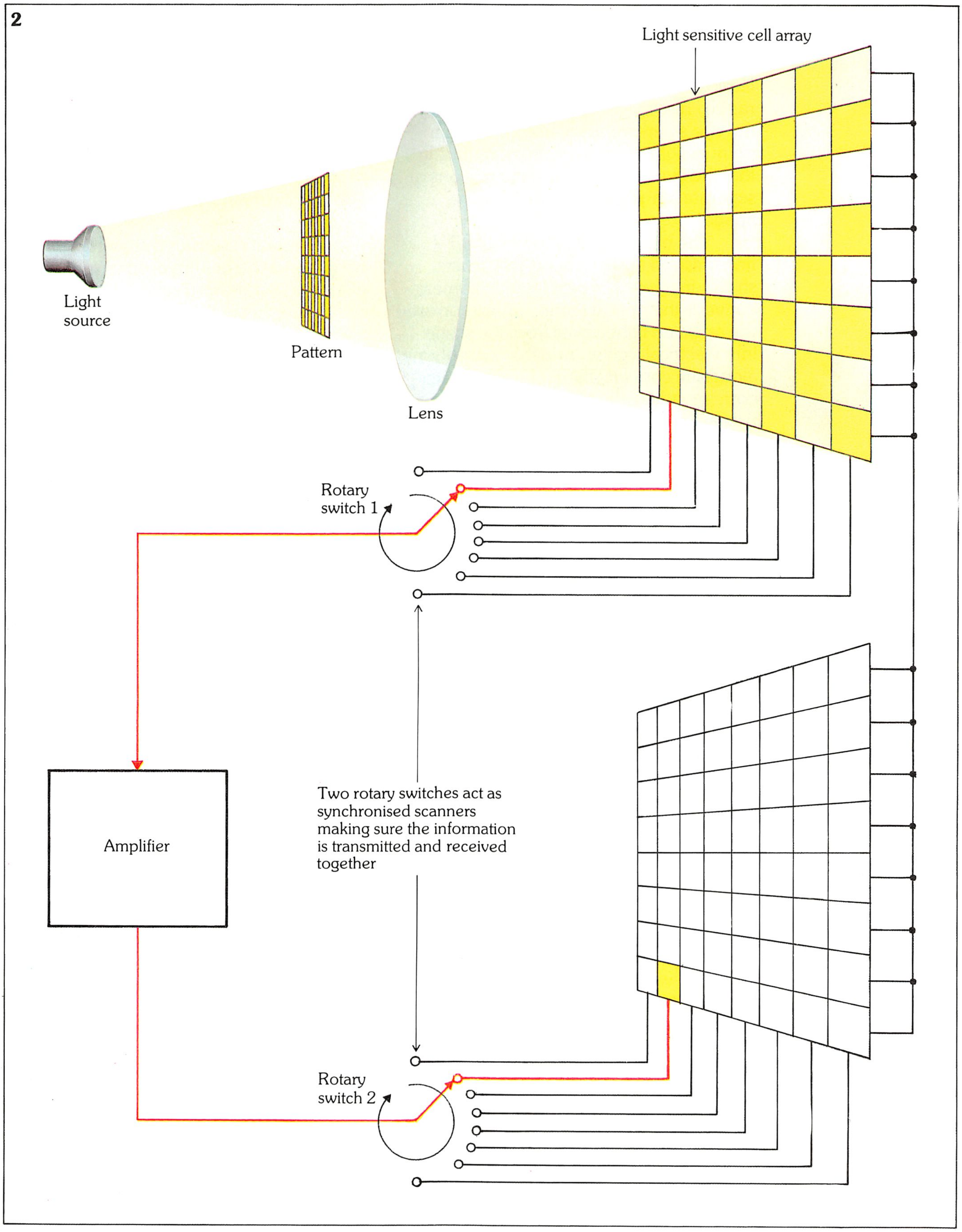
Figure 2 illustrates a system in which two rotary switches are used to switch the electrical signals from cells to bulbs. The rotary switches act as **scanners**. Rotary switch 1, for example, scans each cell in turn and allows the appropriate electrical signal to be transmitted to rotary switch 2. This switch scans at the same rate as the other, and so directs the signal to the corresponding bulb for each cell.

To prevent confusion, and to **synchronise** the two scanners, signals such as 'switch to the second line now', or 'go back to the first line now', could easily be incorporated in the signal stream. For example, a row of 10 'off' signals might instruct the receiving scanner to start again at cell 1; this couldn't be confused with any other signal, as even an all black eight line grid only has eight 'off' signals before a 'new line' signal is received.

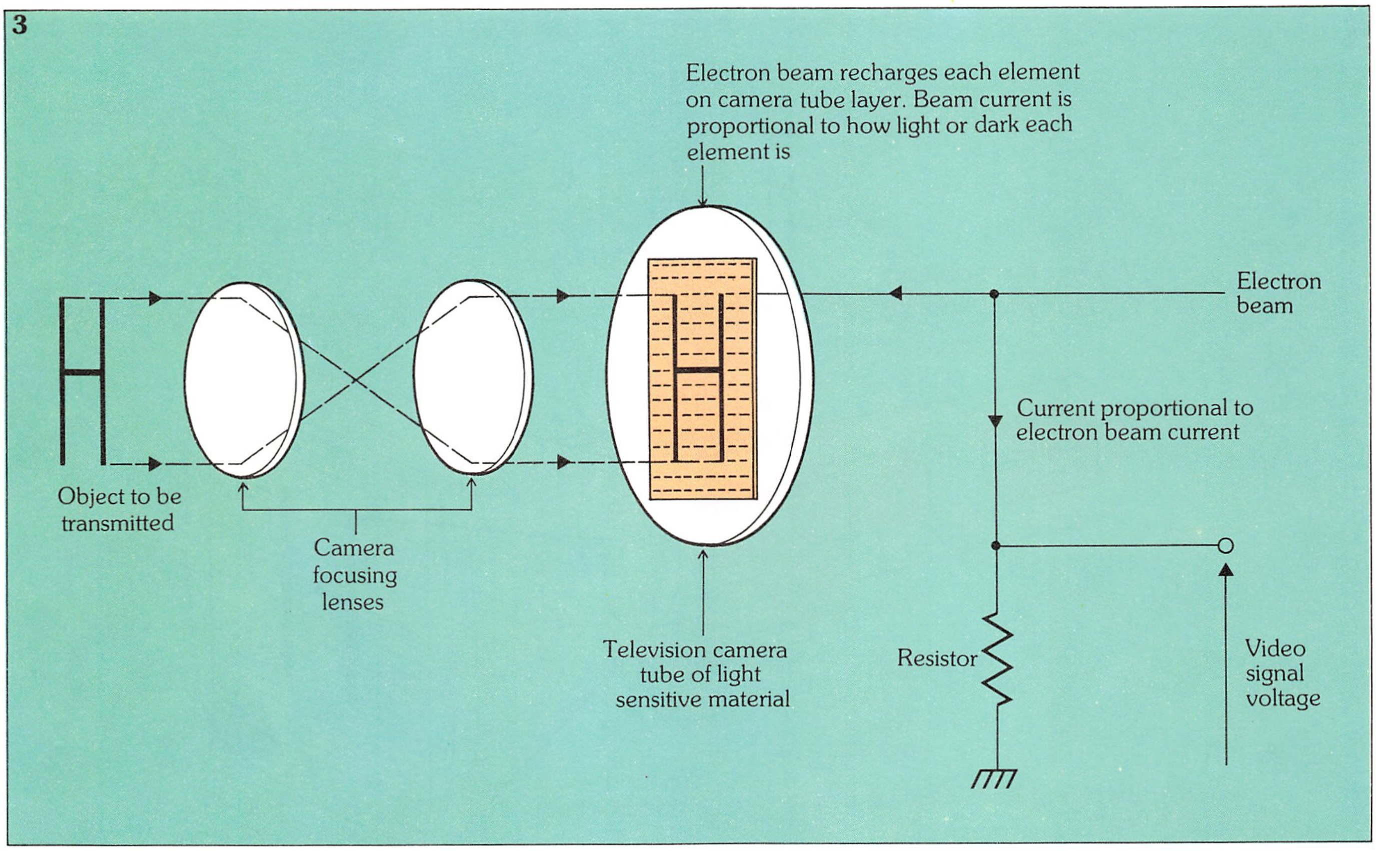
Of course, the mechanical scanning system of *figure 2* is not practicable. However, the principle behind it is. At slow scanning speeds, the image of the chessboard would appear to flash on and off – bulb by bulb. In television terms this is known as **flicker**. As the scanning speed increases, we reach a state where the image presented on the lamps appears continuously lit with a flicker free image of the chessboard. Build the matrix up to 525 or 625 lines of cells and bulbs, instead of the original eight line grid, and the image transmitted begins to gain some definition, rather like a very grainy newspaper illustration.

Scanning by electron

Because mechanical scanning systems lack the speed and reliability necessary for a continuous image, practical television systems scan with the aid of electron beams. At the transmitter, the electron beam reads the illumination levels on the light sensitive


2
Light sensitive cell array
Light source
Pattern
Lens
Rotary switch 1
Amplifier
Two rotary switches act as synchronised scanners making sure the information is transmitted and received together
Rotary switch 2

3



2. Using two rotary switches (acting as synchronised scanners) to switch electrical signals from cells to bulbs.

3. Action of the vidicon tube.

face of a camera tube; at the receiver, an electron beam is used to scan the screen of the picture tube, and display the image transmitted.

These electrons, as we've seen in *Solid State Electronics 23*, are fired from the cathode of an electron gun. Some electrons are only loosely bound to their atoms, so that when the cathode is heated, the electrons gain sufficient energy to break free from the surface of the metal. These free electrons are then concentrated into a beam, which can be bent and directed at a target by electrical or electromagnetic fields provided either by magnetic yokes on the neck of the tube, or electric plates inside. The beam is quickly and easily moved over the face of the tube simply by changing the electrical or electromagnetic fields.

Camera tubes – the vidicon

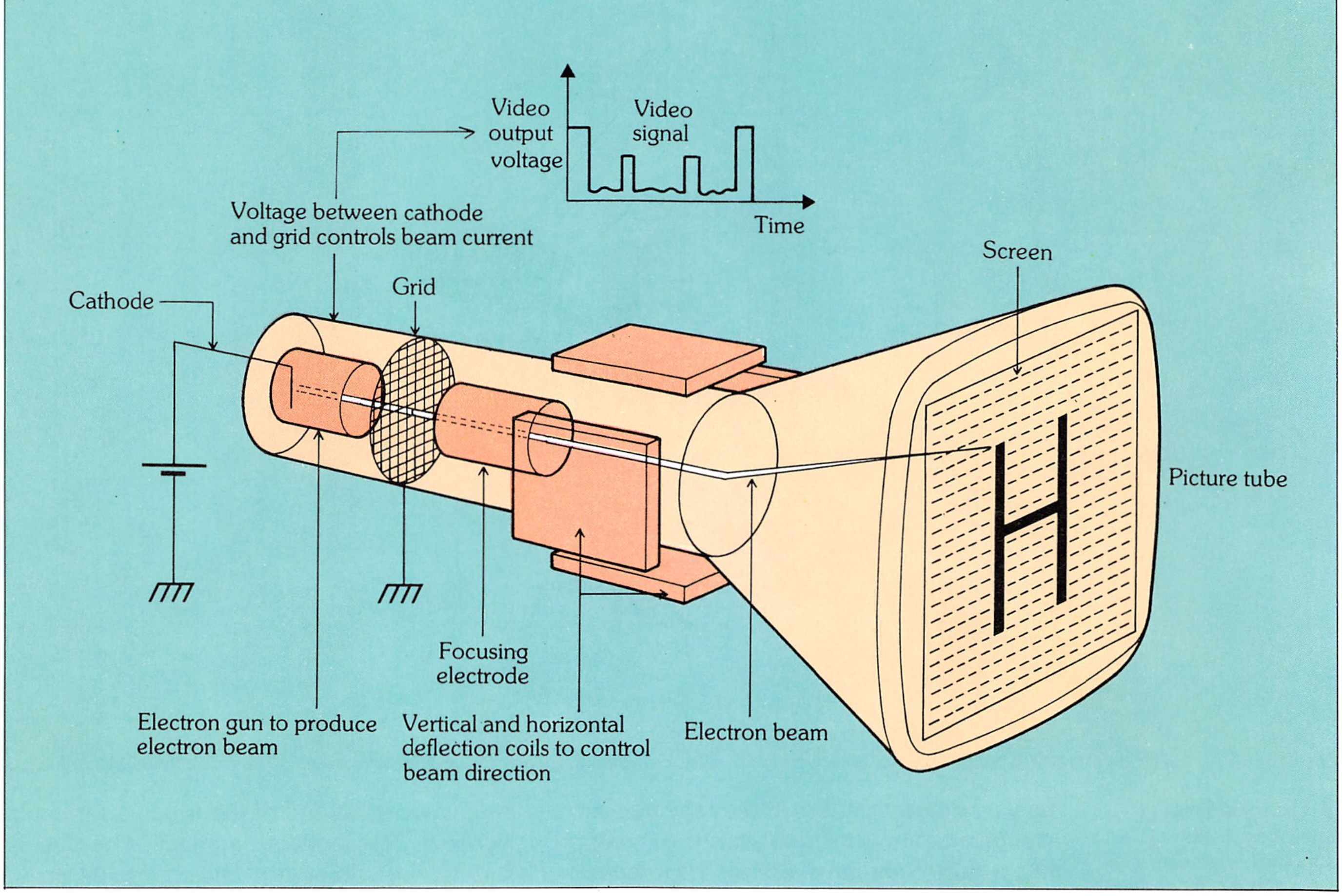
There are many different types of camera tube available but all work on the principles of either photo-conduction or photo-emission. In the **vidicon tube**, the camera lens forms an image of the scene being televised onto a layer of light sensitive material at one end of the tube as shown in *figure 3*. This layer is electrically charged by the electron beam. The charge on an individual element of the layer tends to leak away, i.e. discharge, at a rate dependent on its exposure to light.

If the image on an element is dark, the element loses little charge. The next time the beam sweeps that same element in the photosensitive layer, only a few electrons are needed to replace the lost charge. If the image on the element is light, on the other hand, a considerable amount of charge is lost and the beam must therefore provide a greater number of electrons to recharge the element. The beam current, therefore, is directly proportional to the amount of light on each element.

This current passes through a resistor, so the voltage produced across the resistor is, in turn, also proportional to the amount of light on each element. This voltage is the **vision signal** (more commonly known as the **video signal**) which is transmitted to the television receiver.

The opposite of this process occurs at the receiver. The received video signal

4



controls the intensity and direction of the electron beam. The electrons are focused to produce a fine spot on the picture tube face as shown in *figure 4*. The screen of the picture tube is coated with a phosphor which emits light in proportion to the number of electrons hitting it. So the variations in electron beam intensity produce the same variations in light intensity from the screen, corresponding to the original video signal.

A high voltage signal from the camera indicating a white element allows maximum beam current and provides the greatest light intensity of the phosphor coating. A black element creates a low voltage allowing minimum beam current, and therefore smallest light intensity.

Scan pattern

To form a video signal in the television camera tube, a deflection system scans the electron beam across the image being televised. It scans across the image in horizontal lines from left to right, moving down the screen. When it reaches the bottom, it returns to the top for another scan.

Specific signals are then implanted with the video signal to co-ordinate the movement of the receiver's electron beam with that of the camera. These **synchronising pulses** or **sync pulses** are generated in the television studio and the complete video signal, formed by camera signal and sync pulses, is known as the **composite video signal**. We'll see what this looks like later.

There are two basic types of sync pulses: short duration pulses which control the horizontal scan, from line to line, and are therefore known as **line sync pulses**; and longer duration pulses, **field sync pulses**, which control the vertical scan, from screenful of lines to screenful of lines. (A screenful is known as a **field** of lines.)

(continued in part 33)

4. At the receiver, the video signal controls the intensity and direction of the electron beam producing a fine spot on the picture tube face.